A Metabolic Profiling Approach to Human Disorders of Energy Metabolism

by

Oded Shaham

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN
BIOINFORMATICS AND INTEGRATIVE GENOMICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Author .....
Harvard-MIT Division of Health Sciences and Technology
July 8, 2009

Certified by .....
Vamsi K. Mootha, MD
Associate Professor
Department of Systems Biology, Harvard Medical School
Thesis Supervisor

Accepted by .....
Ram Sasisekharan, PhD
Director, Harvard-MIT Division of Health Sciences and Technology
Edward Hood Taplin Professor of Health Sciences and Technology and Biological Engineering

A Metabolic Profiling Approach

to Human Disorders of Energy Metabolism

by

Oded Shaham

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on July 8, 2009, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY IN
BIOINFORMATICS AND INTEGRATIVE GENOMICS

Abstract

The integrated network of biochemical reactions known collectively as metabolism is essential for life, and dysfunction in parts of this network causes human disease - both rare, inherited disorders and common diseases such as diabetes mellitus. The study of metabolic disease depends upon quantitative methods which are traditionally custom-tailored to a given compound. Recent advances in technologies such as mass spectrometry now enable the simultaneous measurement of a diverse metabolite collection spanning multiple biological pathways, an approach known as metabolic profiling or metabolomics. This dissertation describes the development of one such metabolic profiling system and its application to the study of two major topics in human energy metabolism: the fasting:feeding transition and mitochondrial disease. In the first study, we profile human plasma in response to glucose ingestion, detecting dozens of metabolite changes and identifying several distinct effects of insulin. Based on these observations, we propose a multivariate view of insulin sensitivity, and show that individuals at risk for developing diabetes mellitus can differ in their insulin response profile, a concept of potential value for estimating disease risk and progression. In the second study, we elucidate a metabolic signature of human mitochondrial disease that reflects substrate oxidation, biosynthesis and energy charge. We demonstrate that the culture media profile of a cellular disease model of mitochondrial dysfunction reflects the plasma profile of human patients, an approach that could be applicable to other diseases as well. In addition, we show that a combination of metabolites distinguishes individuals with mitochondrial disease from healthy individuals better than the currently used diagnostic markers. Our findings provide insight into human disorders of energy metabolism, and demonstrate the utility of a profiling approach for the understanding of metabolic disease.

Thesis Supervisor: Vamsi K. Mootha, MD
Title: Associate Professor
Department of Systems Biology, Harvard Medical School

Acknowledgments

I am grateful for support I have received from colleagues and friends, which has made the writing of this dissertation possible.

First, I am thankful to Vamsi Mootha, my research supervisor, for his instruction and advice. Vamsi's enthusiasm and his thoughtful approach to medical research have been a true inspiration. I have been fortunate to have Vamsi as a mentor.

Over the last four and a half years, the Mootha lab has been my daily work environment. While it has changed a lot in size and composition over this period of time, it has always been a friendly and supportive group of people with whom it was fun to discuss research or just hang out. I thank all members of the Mootha lab for creating this unique environment.

I thank David Sabatini and Katherine Sims for serving on my thesis committee and for their advice over the last two years. To Kathie I am further indebted for inviting me to attend her mitochondrial clinic at Massachusetts General Hospital, allowing me to better understand mitochondrial disease.

The study of the human response to a glucose challenge described in this dissertation has begun as a collaboration with the Clinical Research Center at the Massachusetts Institute of Technology. The welcoming approach of the Center's directors and staff has made this collaboration a very pleasant one. I am particularly thankful to Ravi Thadhani, Catherine Ricciardi and Ilene Horvitz for their support.

I am thankful to the Broad Institute of MIT and Harvard for providing resources and funds that supported my research work. I am especially thankful to members of Broad's Metabolite Profiling group for developing technology that was essential for my thesis research.

Finally, I am fortunate to have my family's constant support. My parents Vera and Uzi Shaham and my sisters Nira and Dahlia and their families have been a source of strength and advice. Most of all I thank my wife Sharon for being with me throughout this journey, encouraging me at difficult moments and sharing the good ones.

Contents

Chapter 1 - Introduction 9
Chapter 2 - Development of a Mass Spectrometry-Based Metabolic Profiling System . . . 25
Chapter 3 - Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity 37
Chapter 4 - A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent cell culture media 65
Chapter 5 - Implications and Future Directions 89
Appendix A - Supplementary Material for Chapter 3 97
Appendix B - Supplementary Material for Chapter 4 111

Chapter 1

—

Introduction

Introduction

The term Metabolism stands for the collection of chemical reactions by which living cells extract energy from their environment and convert foodstuffs into cell components (1). Centuries of research, leading to breakthroughs during the 20th century, uncovered fundamental principles and yielded integrated theories concerning the thermodynamics, kinetics, and regulation of metabolism in cells and organisms (1, 2). This knowledge has played an important role in the understanding of human physiology and disease (3). The description of human metabolic disorders was pioneered by Sir Archibald Garrod, who coined the term "inborn error of metabolism" in the beginning of the 20th century. Since then hundreds of such diseases have been described. These rare, inherited disorders, as well as common, complex diseases with a prominent metabolic component, such as diabetes mellitus, are collectively known as metabolic diseases.

The study of metabolism and metabolic disease has always depended on the availability of means to quantify the abundance of molecules. The first description of an inborn error of metabolism, alkaptonuria, was possible thanks to the existence of a method for the quantification of homogentisic acid, a compound whose elevation in urine is the defining feature of this disorder. A variety of chemical methods for quantitative analysis have been developed and applied in biological research, though due to the structural diversity of molecules participating in metabolism each molecule or family of molecules requires a specialized measurement method. Whereas the uniform structure of nucleic acids facilitates the use of generic methods for global analysis, enabling the rapid advancement of genomic technologies and research we see today, the systematic measurement of metabolites is more challenging.

Metabolic Profiling Technology

Recently, technological advances have enabled the development of analytical methods that measure many metabolites simultaneously, creating new opportunities for the study of metabolic disease (4). These advances have also inspired hopes to measure the entire collection of metabolites in an organism, which was termed "metabolome" in analogy to the genome. The two main technologies in this field are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). Methods based on these two technologies can be either targeted to a pre-defined set of metabolites or untargeted, and can provide either absolute or relative quantification. This section describes current metabolic profiling technologies and the types of analytical methods they support.

Nuclear Magnetic Resonance Spectroscopy

In NMR spectroscopy, a magnetic field probes magnetic atomic nuclei within molecules, eliciting a molecule-specific response and providing information on molecular structure and abundance. NMR spectroscopy focusing on the resonance of proton nuclei, known as ^{1}H (proton) NMR spectroscopy, has been adapted to provide quantitative information on a wide spectrum of metabolites. Other magnetic nuclei have also been used and provide distinct advantages: ^{13}C spectra are simpler than proton spectra and do not require solvent suppression (5), while ^{31}P spectroscopy probes phosphate-containing molecules such as ATP, ADP and phosphocreatine, providing information on the biochemical energy status in cells and tissues (1). A key advantage of NMR spectroscopy over mass spectrometry is its non-destructive nature, allowing *in vivo* studies. The main disadvantage of NMR spectroscopy is its low sensitivity compared to mass spectrometry.

Mass Spectrometry

In MS, molecules are ionized and delivered into an electrical field, where they are distinguished based on their mass:charge (m/z) ratio. A key advantage of MS is its high sensitivity. A type of mass spectrometer commonly used for metabolic profiling is the triple quadrupole. A quadrupole consists of 4 circular rods setup to create a well controlled electrical field. A triple quadrupole mass spectrometer, also known as a tandem mass spectrometer (or MS/MS), consists of three quadrupoles arranged in series. The first quadrupole selects ions based on m/z ratio, the second acts as a collision cell where ions are fragmented, and the third selects a fragment ion based on m/z ratio (6). MS/MS machines can simultaneously acquire data on hundreds of precursor-fragment ion pairs, a technique known as multiple reaction monitoring (MRM). In conjunction with MS, chromatography is often used to separate between molecular species in a complex biological sample, delivering them into the mass spectrometer spread out in time. Both gas chromatography and liquid chromatography have been used for metabolic profiling. While gas chromatography requires derivatization of metabolites prior to analysis, liquid chromatography does not impose this restriction.

Targeted versus Untargeted Analysis

Metabolic profiling methods can be divided into targeted methods and untargeted ones. Targeted methods quantify a pre-defined set of identified metabolites, and require the characterization of each targeted metabolite ahead of time. Untargeted methods, on the other hand, acquire entire spectra, from which subsequent data processing methods extract distinct features. The advantage of targeted analysis is that the identity of the measured analytes is known, which enables biological interpretation of data and helps to avoid confounding effects. At the same time, non-targeted methods can potentially detect unexpected metabolites that were not previously considered. Both MS and

NMR spectroscopy can be used in a non-targeted mode, whereas targeted analysis is done only with MS. Targeted MS analysis is accomplished by the combination of chromatography and tandem mass spectrometry with multiple reaction monitoring (6), identifying each metabolite by a unique combination of chromatographic retention time, precursor ion m/z ratio and fragment ion m/z ratio. Mass spectrometry technologies that provide a mass resolution high enough to determine unambiguously the chemical formula of an analyte, such as Fourier transform ion cyclotron resonance (FTICR) or Orbitrap, can identify metabolites in an untargeted manner (6). In NMR spectroscopy metabolite identification is achieved by comparing spectra of biological samples to reference spectra of pure, individual metabolites (7).

Relative vs. Absolute Quantification

The quantification of metabolites can be absolute, meaning that a molar concentration is determined for each metabolite, or relative, meaning that metabolite levels are expressed in arbitrary units. The advantage of absolute quantification is that the results can be compared to other experiment and to published literature, whereas relative quantification only allows comparison to other samples in the same experiment. To obtain absolute quantification, a metabolite needs to be analyzed over a range of known concentrations, creating a standard curve that relates abundance readings to molar concentrations. In mass spectrometry, this is achieved by adding to the biological sample pure metabolite standards containing one or more stable isotope atoms. An isotope-containing metabolite co-elutes with the corresponding endogenous metabolite but its mass/charge ratio is offset by a known number of atomic mass units, allowing clear separation between the two. The MS ionization efficiency is molecule-specific, and as a consequence a separate standard curve is required for each metabolite. Due to the limited availability and high price of isotope-labeled metabolites, and the amount of labor required, it is usually not feasible to obtain absolute quantification for a large number of metabolites. Proton NMR

spectroscopy, on the other hand, is much better suited for absolute quantification since all protons show an identical response factor, regardless of the chemical properties of the molecules containing them (7).

Human Energy Metabolism

I use the term human energy metabolism broadly to refer to the extraction of energy from nutrients, its storage and expenditure, on the whole body level and on the cellular level. In this section I introduce two topics in human energy metabolism. The first is the transition of the human body from a fasting state to a fed state and its relation to insulin sensitivity. The second is the dysfunction of mitochondria, a cellular organelle central to ATP synthesis. While the two topics are discussed separately in this dissertation, it is important to note that mitochondrial function affects whole body energy metabolism. Indeed, previous work has revealed a connection between mitochondrial gene expression and diabetes mellitus (8).

The Fasting-Feeding Transition and Insulin Sensitivity

The transition of the human body from a fasting state to a fed state involves multiple organ systems and metabolic pathways (9). During fast, the body derives energy from food stores, primarily in the form of glycogen and triacylglycerol, and during a prolonged fast also from protein (10). The liver breaks down glycogen and releases glucose, while in adipose tissue triacylglycerol molecules are broken to glycerol and free fatty acids, which are released to the circulation. Fatty acids are further degraded in the liver to produce ketone bodies, which support brain function. When glycogen stores are depleted, alanine derived from muscle protein and lactate derived from glycolysis in various tissues are used in the liver for gluconeogenesis, in processes known as the alanine cycle and the Cori cycle. Ingestion and absorption of food triggers

the release of insulin from the pancreas which switches the body from using up fuel reserves to storing the ingested food as glycogen, lipids and proteins. The body's ability to switch efficiently is important for future survival in periods of starvation, and is dependent on the ability of the pancreas to release insulin and on the sensitivity of the liver, muscle and adipose tissues to the insulin signal.

Diabetes mellitus is a group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both (11). These diseases affect 8% of the population in the United States according to a current estimate by the American Diabetes Association. Type 2 (T2DM) is the most prevalent type of diabetes, characterized primarily by resistance to the action of insulin. Hepatic insulin resistance results in increased release of glucose from the liver, whereas resistance in adipose and muscle tissues causes reduced glucose uptake, all of which contribute to hyperglycemia. During the early stages of disease progression, pancreatic beta cells attempt to compensate for the insulin resistance by secreting increased amounts of the hormone, but after a while their capacity to secrete insulin declines. Chronic diabetes mellitus is associated with severe complications in eyes, kidneys, the nervous system, heart and blood vessels (11), causing major health problems. While intense research efforts aim to elucidate the genetic factors contributing to the development of T2DM, the pathogenesis of this disease is still not well understood.

Loss of sensitivity to the action of insulin is an early sign for the development of T2DM. Insulin sensitivity can be measured directly using the hyperinsulinemic euglycemic clamp technique (12), in which the rate of glucose infusion required to maintain constant blood glucose levels under high insulin conditions reflects the rate of glucose uptake. This technique is difficult to perform in a clinical setting and is limited to research. Simpler indices of insulin sensitivity have been developed and compared to the clamp technique (13). The

fasting insulin level is a simple index providing good correlation with clamp-measured insulin resistance. The homeostasis model assessment (HOMA) of insulin resistance and beta cell function is based on the combination of fasting insulin and fasting glucose concentrations (14). Other indices have been proposed based on time course measurements following glucose ingestion. All indices of insulin sensitivity are based on the effect of insulin on glucose alone. While hyperglycemia is the defining feature of diabetes mellitus, insulin has multiple other physiologic effects, and the action of insulin along these other axes could provide clinically useful information.

Glucose ingestion after overnight fast is a special case of a fasting-feeding transition, where the meal consists of glucose only. This physiologic situation has clinical utility for the diagnosis of diabetes: in the oral glucose tolerance test (OGTT) an individual ingests 75 grams of glucose after overnight fast, and blood glucose is measured two hours post-ingestion. A glucose level higher than 140 mg/dl indicates impaired glucose tolerance and a level exceeding 200 mg/dl indicates diabetes, according to American Diabetes Association guidelines (11). Metabolic profiling of the human response to a glucose challenge can potentially describe the spectrum of metabolic processes involved in the fasting-feeding transition, reveal the multiple physiologic effects of insulin and provide insight into the pathogenesis of insulin resistance and T2DM. In chapter 3 of this dissertation I describe the application of a metabolic profiling system to the OGTT in healthy individuals as well as in individuals with impaired glucose tolerance.

Mitochondrial Dysfunction

Mitochondria play a central role in cellular energy metabolism, performing the final steps in the oxidation of sugars, fatty acids and amino acids, and using the released energy to regenerate ATP in a process known as oxidative phosphorylation (OXPHOS). This process is carried out by the respiratory chain, a series of five enzymatic complexes embedded in the inner mitochondrial

membrane. The respiratory chain is genetically controlled by two genomes: the compact mitochondrial DNA (mtDNA) which encodes 13 protein subunits, and the nucleus which encodes the remaining ~77 subunits. The respiratory chain regenerates ATP by creating a proton gradient across the inner mitochondrial membrane using complexes I through IV, and coupling the dissipation of this gradient to ADP phosphorylation by complex V (15). The proton gradient is created by transporting electrons from the soluble electron carrier NADH and from various oxidation reactions to oxygen, the final electron acceptor, simultaneously with removal of protons from the matrix. In addition to their role in bioenergetics, mitochondria are also involved in other metabolic pathways involving amino acids, ketone bodies and pyrimidine nucleotides. In addition, mitochondria also control key regulatory parameters of the cellular environement including the energy charge and the ratio of [NADH]/[NAD^+], which affect various cellular processes. While it is well recognized that mitochondria have a central role of in cellular metabolism, a systematic description of the metabolic alterations caused by mitochondrial dysfunction has not been possible due to the limitations of traditional analytical methods.

Human mitochondrial disease is broadly defined as a clinically significant functional failure of the mitochondria. More specifically, the term mitochondrial disease commonly refers to a primary defect in the respiratory chain. It is estimated that respiratory chain disease (RCD) affects at least 18.4 in 100,000 people (16) and is the most common group of inborn errors of metabolism (17). Inherited respiratory chain defects are caused by mutations that reside in the mtDNA or in the nucleus. The first mutations were discovered in the mtDNA, following its complete sequencing in 1981 (18). Over 150 different pathogenic mtDNA mutations and many other rearrangements have been discovered in this small genome (19). These mutations and rearrangements affect the function of one or more of the 13 respiratory chain proteins encoded by the mtDNA. Disease-causing mutations in nuclear genes have been found in structural units of the respiratory chain, factors necessary for its maturation and assembly,

genes that support the maintenance of mtDNA and genes that control the lipid composition of the inner mitochondrial membrane (20). Dozens of nuclear genes underlying RCD have been identified, and have enabled genetic testing for these disorders. Our group has been successful in the systematic discovery of such genes through integrative genomics (21-23). While our understanding of the genetic basis of mitochondrial disease has greatly improved in recent years, the dual genetic control over the respiratory chain and the multitude of genes and mutations complicate the molecular diagnosis of this disease.

The diagnosis of RCD is challenging not only due to its genetic basis, but also because of its variable clinical presentation. The clinical features of respiratory chain disease span multiple organ systems including the central and peripheral nervous systems, muscle, eye, blood, endocrine, gastrointestinal, hearing, kidney and liver (24). For example, a relatively common mitochondrial syndrome caused by an mtDNA mutation, Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke (MELAS), typically affects the central and peripheral nervous systems, but can also cause cardiomyopathy, diabetes mellitus, sensorineural hearing loss and intestinal pseudo-obstruction (24). Several existing methods for the measurement of respiratory chain function help support the diagnosis. The enzymatic activity of the entire respiratory chain or individual complexes can be measured in biopsy material, usually muscle. Histochemical staining of complex IV (Cytochrome c oxidase, COX) and complex II (Succinate dehydrogenase, SDH) in a muscle biopsy can reveal low COX activity in the presence of normal SDH activity, indicative of an mtDNA defect. Phosphorous (^{31}P) NMR can be used to measure ATP and phosphocreatine in muscle or brain and to test the rate of ATP regeneration in muscle during recovery from exercise. While these methods provide valuable information, they are either invasive or require special instrumentation or laboratory techniques. Several diagnostic algorithms have been proposed for integrating the clinical, biochemical and molecular findings into a level of certainty in the presence of respiratory chain

disease (25-27), though due to the complexity of these algorithms they are not commonly used in clinical practice.

The abnormal level of metabolites in biological fluids is an important clue for the diagnosis of respiratory chain disease. Metabolites are commonly measured in blood plasma and in urine, and in patients with neurologic symptoms also in cerebrospinal fluid (28). Respiratory chain dysfunction decreases the capacity for pyruvate oxidation, and excess pyruvate can be transaminated to form alanine or reduced to form lactate (29). All three metabolites are often elevated in respiratory chain disease, and have been included in diagnostic algorithms (25-27). Lactate, the most commonly used of the three, is often not elevated in patients with respiratory chain disease, and may be elevated due to various other causes unrelated to RCD (28). Since conversion of pyruvate to lactate by lactate dehydrogenase is coupled to the conversion of NADH to NAD^+, the molar ratio of [lactate]/[pyruvate] is thought to reflect the intracellular [NAD^+]/[NADH] ratio, which is affected by the capability of the respiratory chain to oxidize NADH (30). Based on this rationale, an elevated [lactate]/[pyruvate] ratio in biological fluids is used to differentiate respiratory chain dysfunction from other causes of impaired pyruvate oxidation, such as a pyruvate dehydrogenase dysfunctional. Elevated intermediates of the tricarboxylic acid (TCA) cycle have also been proposed as indicative of RCD (31, 32). The elevation of various other organic acids has been linked to respiratory chain dysfunction, including the dicarboxylic acids adipate, suberate and sebacate, 3-methylglutaconate, ethylmalonate, 2-ethylhydracylate, methylsuccinate, and the glycine-conjugated compounds tiglylglycine, butyrylglycine and isovalerylglycine (31, 32). The association of these metabolites with respiratory chain dysfunction is based on isolated case reports and their diagnostic utility is unknown.

The central role of mitochondria in cellular metabolism suggests that metabolic profiling of mitochondrial dysfunction might identify a diagnostically

useful signature of the disease and help explain its pathogenesis. Chapter 4 describes an investigation of mitochondrial dysfunction in both a cellular model and in RCD patients.

References

1. Stryer, L. (1988) *Biochemistry.* (W. H. Freeman and Company, New York).
2. Newsholme, E. A. & Start, C. (1973) *Regulation in Metabolism.* (John Wiley & Sons, Chichester, New York, Brisbane and Toronto).
3. Scriver, C. R., Beaudet, A. L., Sly, W. S., & Valle, D. (2004) *Metabolic and Molecular Bases of Inherited Disease.* (McGraw-Hill, New York).
4. Fernie, A. R., Trethewey, R. N., Krotzky, A. J., & Willmitzer, L. (2004) Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* 5:763-769.
5. Dunn, W. B., Bailey, N. J., & Johnson, H. E. (2005) Measuring the metabolome: current analytical technologies. *Analyst* 130:606-625.
6. Lu, W., Bennett, B. D., & Rabinowitz, J. D. (2008) Analytical strategies for LC-MS-based targeted metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* 871:236-242.
7. Xu, Q., Sachs, J. R., Wang, T. C., & Schaefer, W. H. (2006) Quantification and identification of components in solution mixtures from 1D proton NMR spectra using singular value decomposition. *Anal Chem* 78:7175-7185.
8. Mootha, V. K., *et al.* (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267-273.
9. Harris, R. A. & Crabb, D. W. (2002) in *Textbook of Biochemistry with Clinical Correlations*, ed. Devlin, T. M. (Wiley-Liss, New York).
10. Cahill, G. F., Jr. (2006) Fuel metabolism in starvation. *Annu Rev Nutr* 26:1-22.
11. Association, A. D. (2007) Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 30:S42-S47.
12. DeFronzo, R. A., Tobin, J. D., & Andres, R. (1979) Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol* 237:E214-223.

13. Hanson, R. L., *et al.* (2000) Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *American journal of epidemiology* 151:190-198.
14. Matthews, D. R., *et al.* (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28:412-419.
15. Nicholls, D. G. & Ferguson, S. J. (2002) *Bioenergetics 3.* (Academic Press).
16. Uusimaa, J., *et al.* (2007) Prevalence, segregation, and phenotype of the mitochondrial DNA 3243A>G mutation in children. *Ann Neurol* 62:278-287.
17. Thorburn, D. R. (2004) Mitochondrial disorders: prevalence, myths and advances. *J Inherit Metab Dis* 27:349-362.
18. Anderson, S., *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
19. DiMauro, S. (2007) Mitochondrial DNA medicine. *Biosci Rep* 27:5-9.
20. DiMauro, S. & Schon, E. A. (2003) Mitochondrial respiratory-chain diseases. *N Engl J Med* 348:2656-2668.
21. Calvo, S., *et al.* (2006) Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat Genet* 38:576-582.
22. Mootha, V. K., *et al.* (2003) Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A* 100:605-610.
23. Pagliarini, D. J., *et al.* (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134:112-123.
24. (2006) *Mitochondrial Medicine.* (Informa Healthcare).
25. Bernier, F. P., *et al.* (2002) Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology* 59:1406-1411.
26. Morava, E., *et al.* (2006) Mitochondrial disease criteria: diagnostic applications in children. *Neurology* 67:1823-1826.
27. Walker, U. A., Collins, S., & Byrne, E. (1996) Respiratory chain encephalomyopathies: a diagnostic classification. *Eur Neurol* 36:260-267.

28. Haas, R. H., *et al.* (2007) Mitochondrial disease: a practical approach for primary care physicians. *Pediatrics* 120:1326-1333.
29. Dimauro, S., Tay, S., & Mancuso, M. (2004) Mitochondrial encephalomyopathies: diagnostic approach. *Ann N Y Acad Sci* 1011:217-231.
30. Robinson, B. H. (1996) Use of fibroblast and lymphoblast cultures for detection of respiratory chain defects. *Methods Enzymol* 264:454-464.
31. Barshop, B. A. (2004) Metabolomic approaches to mitochondrial disease: correlation of urine organic acids. *Mitochondrion* 4:521-527.
32. Shoffner, J. M. (1999) Oxidative phosphorylation disease diagnosis. *Semin Neurol* 19:341-351.
33. Frazier, D. M., *et al.* (2006) The tandem mass spectrometry newborn screening experience in North Carolina: 1997-2005. *J Inherit Metab Dis* 29:76-85.

Chapter 2

—

Development of a Mass Spectrometry-Based Metabolic Profiling System

Development of a Mass Spectrometry-Based Metabolic Profiling System

In 2005 an initiative at the Broad Institute of MIT and Harvard has begun aiming to develop a metabolic profiling capability. This collaborative effort included the selection and purchasing of instruments for liquid chromatography and tandem mass spectrometry, designating the metabolites that will be targeted and creating a method for measuring them, integrating the system with public bioinformatics resources and evaluating the measurement repeatability. This chapter reviews the development of a metabolic profiling capability at Broad Institute.

Instrumentation and Metabolite Selection

The core metabolic profiling technology we chose was tandem mass spectrometry (MS/MS) coupled with high performance liquid chromatography (HPLC), or in short LC-MS/MS. The main considerations for this choice were the high sensitivity of modern mass spectrometers, the high metabolite specificity achievable by the combination of chromatography with molecular fragmentation and the expertise of the Broad Institute proteomics group in mass spectrometry. In this setup, a metabolite is identified by the unique combination of three parameters: chromatographic retention time, the mass/charge ratio of the precursor ion, and the mass/charge ratio of a major fragment ion. The mass spectrometer is operated in a "multiple reaction monitoring" mode to simultaneously quantify all metabolites targeted.

The major instruments comprising this setup (Figure 1) were a triple-quadrupole mass spectrometer (Applied Biosystems 4000 QTRAP), HPLC systems (Agilent 1100) and a robotic auto-sampler (LEAP Technologies). To

span a wide biochemical spectrum, we used three different chromatography columns (provided by Phenomenex) each geared towards a different broad class of metabolites: Luna Phenyl-Hexyl operating in reversed-phase for amino acids and amine compounds, Luna Amino operating in normal phase for sugars and negatively charged compounds such as nucleotides, and Synergi Polar-RP operating in reversed-phase for polar and acidic compounds such as organic acids. Additional details on the HPLC and mass spectrometry settings can be found in Appendix A.

We developed the metabolic profiling system as a multi-purpose resource for use in various studies of human physiology and disease. Consequently, the metabolites we targeted for monitoring spanned primary pathways of human metabolism, including central carbon metabolism (glycolysis, the pentose-phosphate pathway and the tricarboxylic acid cycle), amino acids and nucleotides. Lipids and fatty acids were slated for future extension of the platform, as they require separate extraction and chromatography methods. In its initial phase, the system was set up to monitor approximately 200 different metabolites (a complete list can be found in Appendix A). For each targeted metabolite we purchased a pure standard, tested it on each of the three columns to select the one providing the best retention and elution parameters, tuned the fragmentation conditions and identified a major fragment to be monitored.

Computational Integration with Public Resources

A long tradition of biochemistry has created a large body of knowledge of metabolism scattered in thousands of books and research papers. Efforts for curation of this knowledge in digital, structured databases (1-4) enable systematic exploration of relationships between genes, enzymes, reactions and metabolites. We have interfaced the metabolic profiling system with several of these databases in order to view our data in the global context of metabolism.

The Kyoto Encyclopedia of Genes and Genomes (KEGG (2)) is a public database linking between genes, enzymes, reactions, pathways and metabolites across multiple organisms. We have stored this information in a relational database and mapped the metabolic profiling system to the database. This database enabled us to estimate how well we cover the various areas of human metabolism (Table 1), which helped in planning the inclusion of additional metabolites. To visualize the coverage of particular biochemical pathways, we highlighted targeted metabolites on individual KEGG pathway maps (Figure 2). We have also used this method to visualize experiment results, which assisted in their interpretation. The metabolic database has also enabled various systematic analyses such as listing all reactions involving a given metabolite, or all reactions linking between certain substrate and product metabolites.

With the development of metabolic profiling technologies, metabolite-centric databases have emerged that provide biochemical information as well as technical parameters for measurement of metabolites. One such database is the Human Metabolome Database (HMDB (4)) which aims to document all human metabolites. We used the biochemical taxonomy of the HMDB to examine the distribution of metabolites targeted by our system into chemical classes (Figure 3). We have also used the published mass spectra of metabolites included in the HMDB to propose candidates for unidentified species.

Repeatability of Metabolite Measurement

In the development of analytical methods it is important to estimate linearity, limits of detection and repeatability (variation). This section describes the estimation of measurement variation in our metabolic profiling system. Estimating the variation within a data sample facilitates proper statistical design of studies that enables detection of meaningful biological effects. Having an

estimate of the variation and knowledge of the expected effect size, one can use a formal statistical power calculation to determine the number of independent observations required for a good chance to detect an effect ((5), Figure 4). This principle applies both to a single analytical assay and to the simultaneous measurement of hundreds of parameters, as in our metabolic profiling system.

The process of profiling a biological sample encompasses sample preparation, LC-MS/MS analysis and peak area integration, all of which can potentially introduce variation. Sample preparation includes multiple steps: protein precipitation using ethanol, drying of the supernatant and reconstitution in water for LC-MS/MS analysis (the protocols are detailed in chapters 3 and 4). To estimate the repeatability of this entire process, we split a single biological sample into several parts and processed each part separately. We performed this experiment on the main two types of biological samples we initially targeted - human plasma and cell culture media. Media was sampled from cultures of C2C12 mouse myoblasts after 24 hours incubation, allowing the cells time to effect a change on the media. We calculated the coefficient of variation (CV, the standard deviation divided by the mean) for each metabolite separately, and summarized the CV distributions (Figure 5). Metabolites targeted by our system span a wide range of concentrations in human plasma and in culture media, from millimolars to micromolars and below. As expected, metabolites with a stronger signal tended to have lower CVs. The median CV in plasma was 19%, whereas in culture media the median CV was only 12%. This difference could be due to the higher mixture complexity of plasma compared to media. The CV distribution of media sampled from separate culture plates was similar to the distribution of media sampled from the same plate (Figure 5 panels B and C), indicating that C2C12 cultures grown under the same conditions have a very similar metabolic profile.

Figure 1: The metabolic profiling system.

The setup consists of 3 HPLC systems (Agilent 1100) controlling 3 different chromatographic columns connected to a mass spectrometer (Applied Biosystems 4000 QTRAP). A robotic arm (LEAP Technologies) automatically samples from 96-well plates.

Metabolism Area	Targeted	Total	Coverage
Amino Acid Metabolism	79	370	21%
Nucleotide Metabolism	50	109	46%
Carbohydrate Metabolism	49	236	21%
Energy Metabolism	35	80	44%
Metabolism of Cofactors and Vitamins	33	148	22%
Metabolism of Other Amino Acids	31	105	30%
Biosynthesis of Secondary Metabolites	23	114	20%
Lipid Metabolism	21	284	7%
Biodegradation of Xenobiotics	12	180	7%
Biosynthesis of Polyketides and Nonribosomal Peptides	6	18	33%
Glycan Biosynthesis and Metabolism	4	90	4%

Table 1: Coverage of human metabolites by metabolic profiling

Coverage of human metabolites in the 11 metabolism areas of the KEGG database (2). Human metabolites were determined based on association with an enzyme coded by a documented human gene.

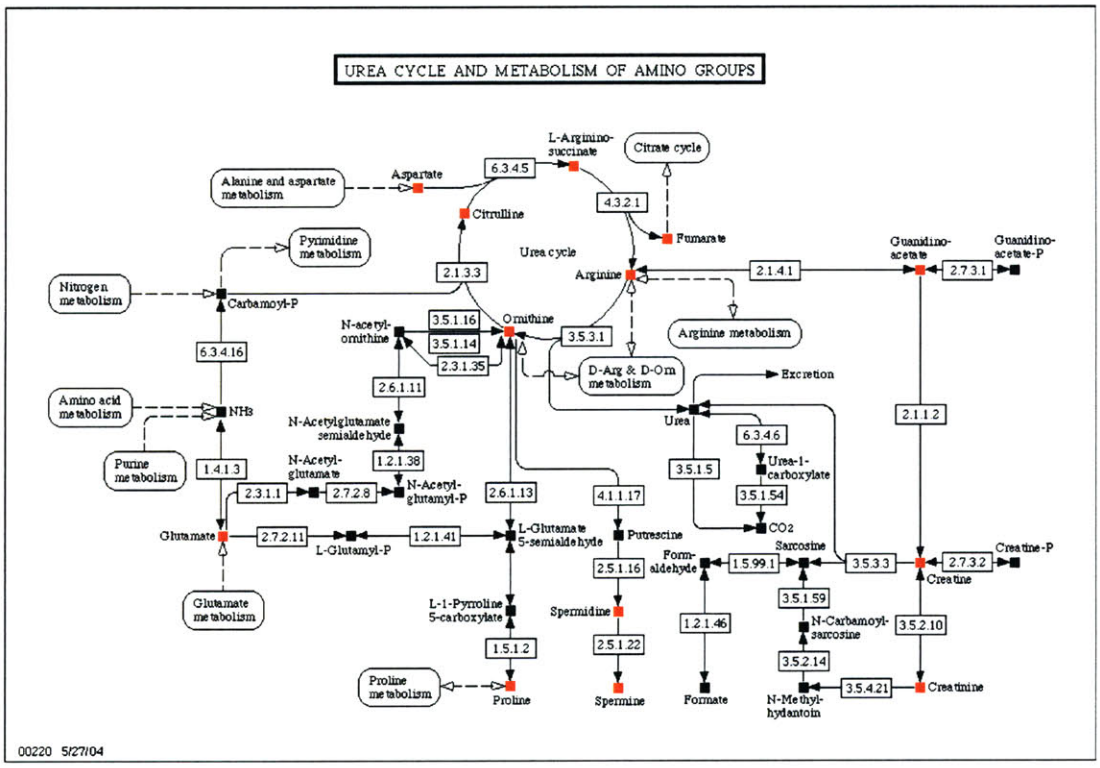


Figure 2: Visualization of metabolite coverage on biochemical pathway maps.

In this example, the KEGG (2) pathway map for urea cycle and metabolism of amino groups is shown. Red squares mark metabolites targeted by the metabolic profiling system.

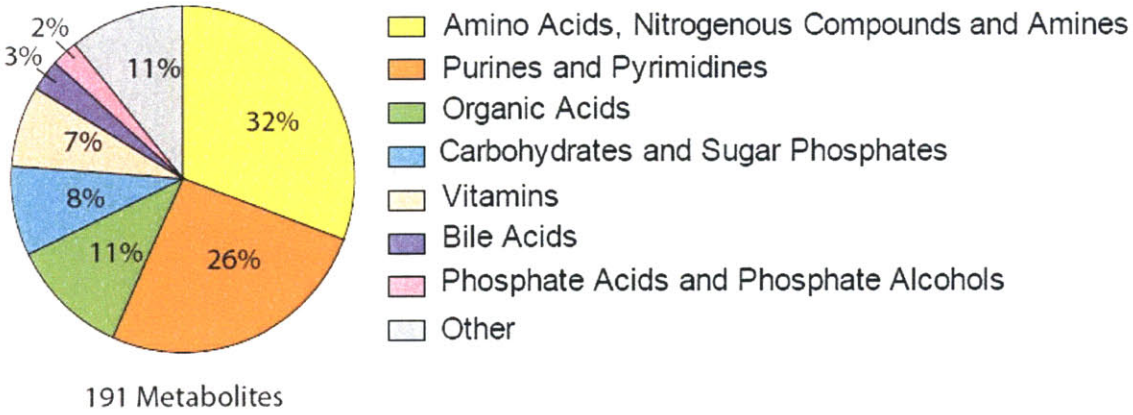


Figure 3: Distribution of targeted metabolites into chemical classes.

The chemical classification is based on the chemical taxonomy annotation in the Human Metabolome Database (4).

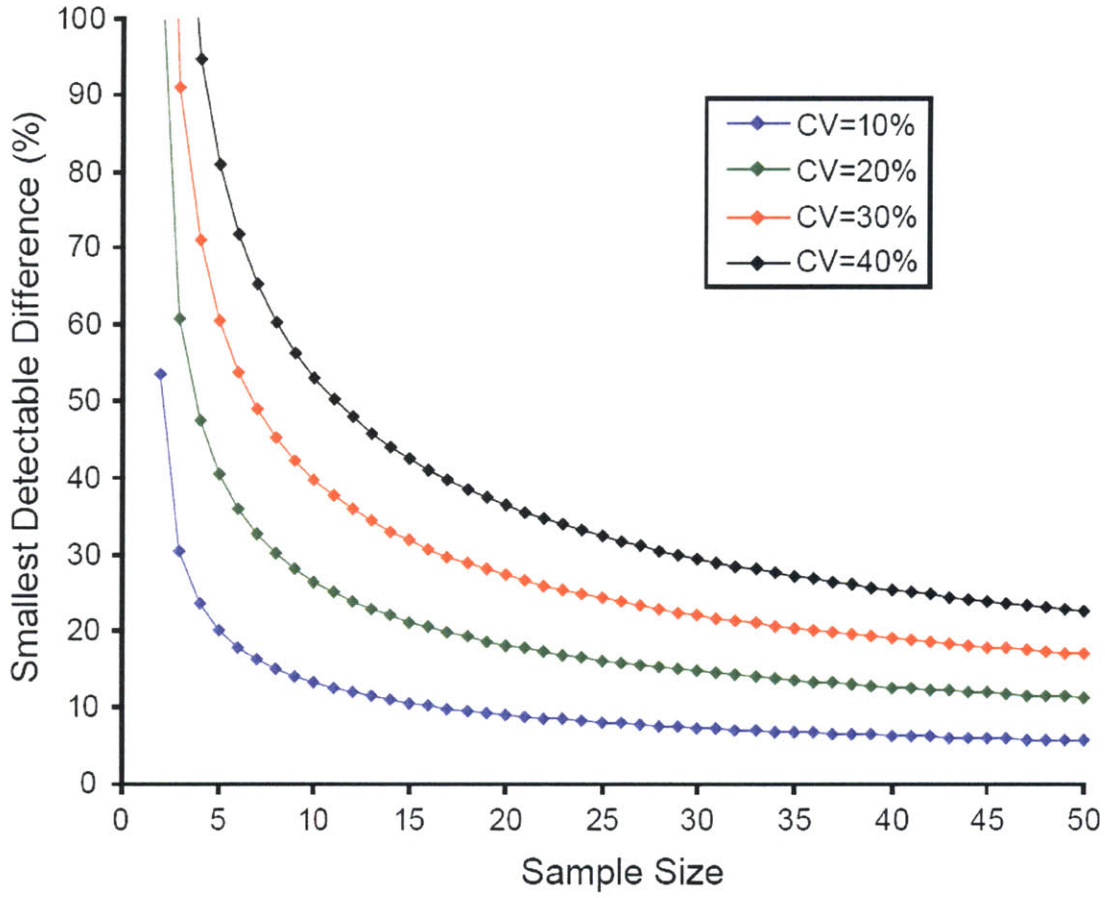


Figure 4: Theoretical calculation of the smallest detectable difference between two samples as a function of the sample size and the coefficient of variation (CV).

Based on Student's t-test (assuming two normally-distributed samples) at P=0.05 and a power of 80% (5).

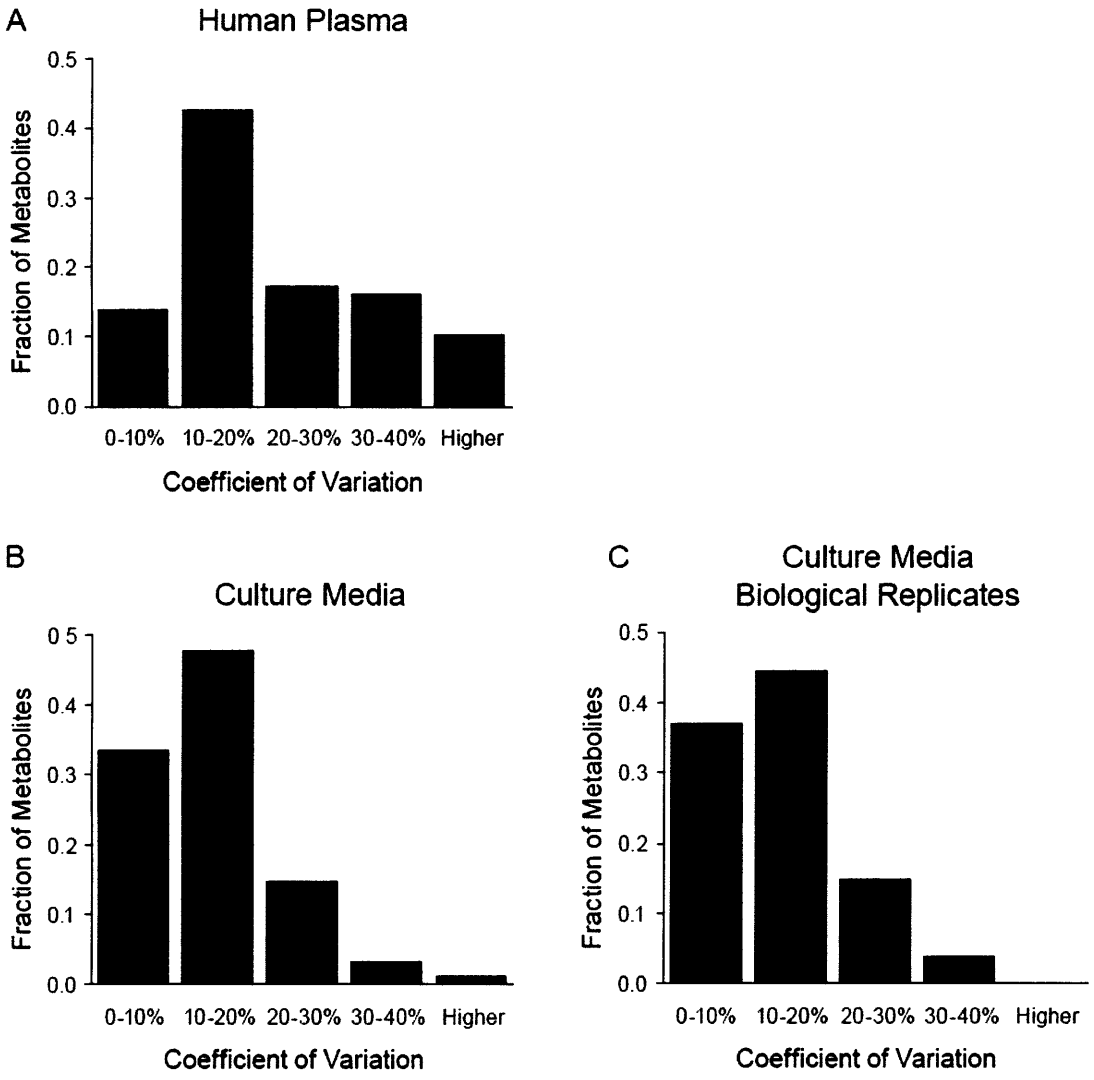


Figure 5: Reproducibility of metabolite measurement.

The variation associated with sample preparation and LC-MS/MS analysis was calculated by splitting one sample into several parts and processing each part separately. The coefficient of variation (the standard deviation divided by the mean) was calculated for each metabolite detected.

A) Human plasma (n=7).

B) Culture media sampled after 24 hour incubation of C2C12 myoblasts (n=8).

C) Culture media as in panel B, except that samples originated from separate culture plates (n=8).

References

1. Duarte, N. C., *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104:1777-1782.
2. Kanehisa, M. & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27-30.
3. Ma, H., *et al.* (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 3:135.
4. Wishart, D. S., *et al.* (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Res* 35:D521-526.
5. Zar, J. H. (1999) *Biostatistical analysis.* (Prentice-Hall, Inc., Upper Saddle River).

Chapter 3

—

Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity

Oded Shaham, Ru Wei, Thomas J. Wang, Catherine Ricciardi, Gregory D. Lewis, Ramachandran S. Vasan, Steven A. Carr, Ravi Thadhani, Robert E. Gerszten and Vamsi K. Mootha

This chapter originally appeared in *Molecular Systems Biology* 4:214 (2008).

Corresponding Supplementary Material can be found in Appendix A.

Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity

Glucose ingestion after an overnight fast triggers an insulin-dependent, homeostatic program that is altered in diabetes. The full spectrum of biochemical changes associated with this transition is currently unknown. We have developed a mass spectrometry-based strategy to simultaneously measure 191 metabolites following glucose ingestion. In two groups of healthy individuals (n=22 and 25), 18 plasma metabolites changed reproducibly, including bile acids, urea cycle intermediates, and purine degradation products, none of which were previously linked to glucose homeostasis. The metabolite dynamics also revealed insulin's known actions along four key axes - proteolysis, lipolysis, ketogenesis, and glycolysis - reflecting a switch from catabolism to anabolism. In pre-diabetics (n=25), we observed a blunted response in all four axes that correlated with insulin resistance. Multivariate analysis revealed that declines in glycerol and leucine/isoleucine (markers of lipolysis and proteolysis, respectively) jointly provide the strongest predictor of insulin sensitivity. This observation indicates that some humans are selectively resistant to insulin's suppression of proteolysis, whereas others, to insulin's suppression of lipolysis. Our findings lay the groundwork for using metabolic profiling to define an individual's 'insulin response profile', which could have value in predicting diabetes, its complications, and in guiding therapy.

Introduction

Glucose homeostasis is a complex physiologic process involving the orchestration of multiple organ systems. During an overnight fast, for instance, glucose levels are maintained through both glycogenolysis and gluconeogenesis, and the liver begins to generate ketone bodies (e.g., acetoacetate, β-hydroxybutyrate) from fatty acids released by adipose tissue (Cahill, 2006). Ingestion of glucose after overnight fasting then triggers the rapid release of insulin from the pancreas, which promotes glucose uptake in peripheral tissues and switches the body from catabolism to anabolism. For example, proteolysis in skeletal muscle and associated release of amino acids (which support gluconeogenesis) are replaced by amino acid uptake and protein synthesis (Felig, 1975; Fukagawa et al., 1985). Also, triacylglycerol lysis in adipose tissue and hepatic synthesis of ketone bodies are inhibited and replaced by fatty acid uptake and re-esterification. Hence the transition from fasting to feeding is accompanied by many changes in metabolite concentrations as the body makes adjustments to achieve glucose homeostasis.

To obtain a systematic view of the physiologic response to glucose ingestion in health and disease, we wished to simultaneously monitor the concentrations of a large and diverse set of metabolites. Such an integrated analysis can aid in the classification of metabolic states, reveal new pathways, and potentially improve the sensitivity for detection of abnormalities. Traditionally, single metabolites or classes of small molecules have been detected using dedicated analytical assays. With these methods, relationships between diverse metabolites and pathways have been missed, and a comprehensive picture of a complex physiologic program has not been possible. Such relationships could, in principle, be important in understanding disease pathogenesis and in aiding in the diagnosis of disease.

Recent advances in analytical chemistry and computation now enable the simultaneous measurement of a large number of metabolites. Two technologies are commonly used for metabolic profiling: nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) (Dunn et al., 2005). While NMR spectroscopy has a number of advantages, primarily its non-destructive nature, its ability to provide information on chemical structure and its better suitability for absolute quantification, it tends to have low sensitivity. MS technology, on the other hand, affords sensitive and specific analysis of metabolites in complex biological samples, particularly when implemented as tandem mass spectrometry and coupled with high performance liquid chromatography (LC), a combination termed LC-MS/MS. Metabolic profiling with LC-MS/MS technology has already been successfully used for identifying human plasma markers of myocardial ischemia (Sabatine et al., 2005) as well as for characterizing the metabolic response to starvation in model organisms (Brauer et al., 2006).

We have developed an LC-MS/MS metabolic profiling system capable of quantifying metabolites in plasma (see Methods), and we report here its application to studying the human response to an oral glucose load. Our system can measure 191 endogenous human metabolites spanning diverse chemical classes (Supplementary data 1). This collection of metabolites includes those previously studied in the context of glucose homeostasis, as well as many metabolites not previously linked to this area. We first applied this technology to healthy individuals in order to characterize the normal human response to an oral glucose challenge, and then asked how this response is affected by reduced insulin sensitivity in individuals with impaired glucose tolerance.

Results

Eighteen Plasma Metabolites Change Significantly and Reproducibly during an Oral Glucose Challenge

To systematically characterize the normal biochemical response to glucose ingestion in humans, we obtained plasma samples for metabolic profiling from an ongoing study, Metabolic Abnormalities in College Students (MACS, see Methods). In this study, young adults undergo a series of metabolic evaluations, including a questionnaire for metabolic syndrome risk factors, indirect calorimetry, measurement of body composition and a fasting blood lipid profile. As part of the metabolic assessment, MACS subjects also undergo a 2-hour oral glucose tolerance test (OGTT) with multiple blood draws. To control for the fasting condition and for the effects of fluid ingestion, a subset of MACS subjects selected at random were given an identical volume of water instead of the glucose solution. Venous blood was drawn during fasting and then every 30 minutes following glucose or water ingestion for the 2-hour duration of the test. We obtained samples from 22 subjects ingesting glucose and 7 control subjects ingesting water (Table 1). Serum concentrations of glucose and insulin were measured throughout the test (Figure 1A). All subjects had normal fasting glucose levels, and all glucose-ingesting subjects showed normal glucose tolerance, as currently defined by the American Diabetes Association (American Diabetes Association, 2007).

We performed LC-MS/MS metabolic profiling of the OGTT time course in MACS subjects. Out of the 191 metabolites monitored by our platform, 97 were detected in at least 80% of subjects in all time points (Figure 1B). The levels of 21 metabolites changed significantly ($p<0.001$) from the fasting levels and were also significantly ($p<0.05$) different when compared to the response to water (Figure 1C). These 21 significantly altered metabolites span pathways previously

studied in the context of glucose homeostasis, as well as some never linked to this program.

To validate these findings, we profiled fasting and 2-hr OGTT samples from an independent group. The 2-hr time point has clinical significance, and can aid in the diagnosis of impaired glucose tolerance and diabetes (American Diabetes Association, 2007). FOS-NGT (Table 1) is a group of individuals with normal glucose tolerance, derived from the Framingham Offspring Study (Kannel et al., 1979). This group is similar to MACS in size, gender composition and glucose tolerance, but these individuals are approximately 20 years older and differ in their ancestry (Table 1). Of the 21 metabolites displaying significant change in MACS at any time-point during OGTT (Figure 1C), the levels of 20 (glucose excluded) remained significantly ($p<0.05$) altered at the 2-hour time point. 18 of these 20 metabolite changes replicated significantly ($p<0.05$) and in the same direction in FOS-NGT (Figure 2). The remaining two metabolites, malate and arginine, fell below our significance threshold. We have thus identified 18 plasma metabolites exhibiting highly reproducible and likely robust responses to glucose ingestion in healthy individuals.

Metabolic Profiling Reveals Novel Biochemical Changes During the OGTT

The systematic profiling approach has enabled us to identify a number of plasma metabolites, not previously associated with glucose homeostasis, that change reproducibly in response to an oral glucose challenge. Perhaps most striking were the observed changes in bile acids. The levels of three bile acids – glycocholic acid, glycochenodeoxycholic acid and taurochenodeoxycholic acid – more than doubled during the first 30 minutes after glucose ingestion, and remained elevated for the entire two hours (Figure 3A). Water ingestion produced a smaller increase in bile acids which was not sustained beyond the 30 minutes time point. All three compounds are primary bile acids conjugated to glycine or taurine.

The levels of citrulline and ornithine, two non-proteinogenic amino acids which participate in hepatic urea synthesis, decreased by 35% and 29% respectively during the 2-hour test (Figure 3B). Gluconeogenesis from amino acids (primarily alanine), which supports 25-40% of the non glycogen-derived hepatic glucose output after an overnight fast (Felig, 1975), is coupled to urea synthesis (Devlin, 2002). The decreases in citrulline and ornithine may thus be associated with the reduction in gluconeogenesis and urea synthesis following glucose ingestion.

Hypoxanthine, a purine base generated from degradation of adenine and guanine nucleotides, decreased in MACS by 39% within two hours of glucose ingestion (Figure 3C), and this pattern was replicated in FOS-NGT. Xanthine, a purine base generated from hypoxanthine by oxidation, also decreased in both groups (MACS: -9%, $p<0.05$; FOS-NGT: -41%, $p<10^{-4}$). The decreases in hypoxanthine and xanthine may be explained by a combination of attenuated release and accelerated uptake. Hypoxanthine taken up by tissues can support nucleotide biosynthesis through the purine salvage pathway (Mateos et al., 1987; Yamaoka et al., 1997) and may also be indicative of a switch from catabolism to anabolism of nucleic acids, analogous to the simultaneous transitions in fat and protein metabolism.

Interestingly, hippuric acid increased by over 1000% during the first 30 minutes and decreased gradually thereafter. Most likely this response is not related to glucose, but rather reflects the presence of the preservative benzoic acid, a precursor of hippuric acid (Kubota and Ishizaki, 1991), found in the glucose beverage used for OGTT (see Methods).

Changes in Plasma Metabolites Span Four Arms of Insulin Action

Much of the biochemical response to glucose ingestion, which we studied in an unbiased way, can be attributed to the broad actions of insulin (Figure 4A). Specifically, we have detected an increase in lactate and decreases in glycerol, β-hydroxybutyrate and amino acids. These metabolite changes correspond to the stimulation of glucose metabolism and to the suppression of lipolysis, ketogenesis and proteolysis (as well as stimulation of amino acid uptake), all of which are known to be elicited by insulin (Fukagawa et al., 1985; Nurjhan et al., 1986).

Our method captured the temporal relationship between glucose and intermediates of glycolysis (Figure 4B). Specifically, the increase of pyruvate, lactate and alanine occurred between 30 and 60 minutes, lagging ~30 minutes behind the glucose rise, consistent with previous reports (Kelley et al., 1988). Interestingly, the kinetics of malate, an intermediate in the Krebs cycle, closely resembled the kinetics of lactate and pyruvate, suggesting that a fraction of the pyruvate formed through glycolysis was carboxylated to generate malate via oxaloacetate. To our knowledge, elevation of plasma malate levels due to glucose ingestion has not been previously reported.

To gain insight into the kinetics of insulin action, we compared temporal patterns for metabolites indicative of the suppression of fat and protein catabolism. The levels of glycerol, β-hydroxybutyrate, and multiple amino acids all declined after glucose ingestion, but the kinetic pattern of glycerol and β-hydroxybutyrate was remarkably different from amino acids (Figure 4C). Over the two hours, the decrease of glycerol and β-hydroxybutyrate levels was 57% and 55% respectively, while the drop in amino acids was moderate (between 14-36%). The branched-chain amino acids leucine/isoleucine (indistinguishable by our method), for example, decreased 33%. Interestingly, the time to reach half-maximal decrease in the amino acids (50-72 minutes) was greater than in β-

hydroxybutyrate (42 minutes) and glycerol (30 minutes). Moreover, the inter-individual variance in metabolite levels shrunk dramatically over the two hours in glycerol and β-hydroxybutyrate (84% and 95% reduction of inter-quartile range, respectively), while in amino acids the maximal reduction was 53%. These findings suggest that the suppression of lipolysis and ketogenesis is more sensitive to the action of insulin compared to suppression of protein catabolism.

Metabolite Excursions Reflect Multiple Dimensions of Insulin Sensitivity

Having identified 18 metabolites that exhibit robust 2-hr excursions, we were interested in determining whether they might be useful in understanding insulin sensitivity. Insulin sensitivity is traditionally defined as the ability of insulin to promote the uptake of glucose into peripheral tissues such as skeletal muscle and fat or to suppress gluconeogenesis in the liver. A decline of insulin sensitivity is one of the earliest signs of type 2 diabetes mellitus (T2DM). This decline is often manifest as elevated levels of fasting insulin, which is strongly correlated to measurements of insulin sensitivity (Hanson et al., 2000). Considering that several metabolic processes taking place following glucose ingestion are mediated by insulin, we hypothesized that insulin sensitivity could be reflected not only by changes in glucose, but also by the OGTT-response of multiple other metabolites. Because our initial studies were focused on normal, healthy individuals spanning a narrow range of fasting insulin levels, we performed a third analysis on a group of individuals with impaired glucose tolerance from the Framingham Offspring Study, FOS-IGT, who spanned a broader range of fasting insulin concentrations (Table 1).

First, to systematically evaluate the relationship between individual metabolite excursions and fasting insulin, we performed linear regression of the fasting insulin concentration on each of the 18 2-hr excursions. Out of the 18, six showed a statistically significant ($p<0.05$) correlation to fasting insulin, and included the excursions in lactate, β-hydroxybutyrate, amino acids

(leucine/isoleucine, valine and methionine) and a bile acid (GCDCA) (Table 2). Taken together with the glycerol excursion, which scored (p=0.07) slightly below the significance threshold, the response of four distinct insulin action markers correlated with fasting insulin (Figure 5A). Individuals with high fasting insulin exhibited a blunted excursion in all seven metabolites: they had a smaller change both in increasing metabolites (lactate and GCDCA) and in decreasing metabolites (the other five). Notably, the glucose excursion was not correlated with fasting insulin (p=0.20). These findings suggest that resistance to the action of insulin on the metabolism of glucose, fat, and protein is reflected by the metabolite response to OGTT.

Next, we sought to determine whether a combination of two or more of the 18 metabolite excursions might be more predictive of insulin sensitivity than are individual excursions. We used forward stepwise linear regression to discover an optimal linear model relating 2-hr metabolite changes to fasting insulin levels. The top regression model identified consisted of the excursions in Leu/Ile and glycerol (R^2_{adj}=0.54, p=0.0001). In this bivariate model, the independent contribution from each metabolite excursion was significant (Leu/Ile: $p=8\times10^{-5}$, glycerol: $p=4\times10^{-3}$), and the two predictors were not correlated with each other (p=0.6). The Leu/Ile-glycerol model predicted fasting insulin levels better than any individual metabolite excursion (Table 2). BMI, which is known to be a strong predictor of fasting insulin, was less predictive than the bivariate model. Notably, the explanatory power of the Leu/Ile and glycerol excursions was significant even after controlling for BMI ($p=2\times10^{-3}$, $p=3\times10^{-3}$ respectively). A graphical representation of the Leu/Ile-glycerol model (Figure 5B) demonstrates that some individuals with high fasting insulin exhibit a blunted decline in glycerol, while others exhibit a blunted decline in Leu/Ile.

To identify linear combinations of all 18 metabolite excursions that might be predictive of fasting insulin, we used partial least squares (PLS). PLS is a multivariate method that finds orthogonal linear combinations (called

components) of the original predictor variables (here, metabolite excursions), which are most correlated with a response variable (here, fasting insulin). To determine how many components should be included in the model while keeping it general, the model is tested on new data (a method called cross-validation). In the current PLS model, the smallest prediction error was achieved with the first component alone (Table 2), and a similar error was obtained with the first two components. Interestingly, the Leu/Ile excursion had the largest coefficient in the first component, while the glycerol excursion contributed primarily to the second. The prediction error of the bivariate model consisting of the excursions in Leu/Ile and glycerol was lower than the PLS model error (Table 2). These findings indicate that the excursions in Leu/Ile and glycerol are sufficient to capture the correlation of 2-hr metabolite changes with fasting insulin.

In summary, we have used a univariate approach as well as two multivariate approaches to investigate the correlates of insulin sensitivity. Remarkably, all three methods spotlighted the change in Leu/Ile and change in glycerol as being predictive of fasting insulin. Changes in these two metabolites are blunted in insulin resistant individuals during the OGTT.

Discussion

In the current study, we have applied metabolic profiling to investigate the kinetics of human plasma biochemicals in response to an oral glucose challenge. To our knowledge, this is the first study to apply a profiling technology to characterize this physiologic program. The systematic approach we have taken has enabled us to recapitulate virtually all known polar metabolite changes associated with the OGTT, while spotlighting some pathways never linked to this program. Importantly, simultaneous measurement of multiple metabolites made it possible to explore connections between metabolic pathways, providing novel insights into normal physiology and disease.

One of the most surprising findings of this investigation is the dramatic increase in bile acid levels following glucose ingestion. In response to eating, the gallbladder contracts and releases bile into the intestines. Bile acids are then absorbed and travel through the portal vein back to the liver, where the enterohepatic cycle completes. While the uptake of bile acids to the liver is fairly efficient, a constant fraction (10-30%) reaches the systemic circulation (Angelin et al., 1982). In a previous investigation, human subjects ingesting a standard liquid meal containing fat and protein exhibited an increase of 320-330% in individual serum bile acids (De Barros et al., 1982). In the current study, we found that ingesting glucose alone elicits a bile acid response of similar magnitude (Figure 3A), which is sustained for two hours. Our finding is supported by previous work showing that glucose ingestion can increase the plasma concentration of cholecystokinin, a hormone signaling the gallbladder to contract (Liddle et al., 1985). Interestingly, a recent investigation has shown that bile acids are a signal for fat and muscle cells to increase their energy expenditure through activation of thyroid hormone (Watanabe et al., 2006). Given these findings, and our observation of bile acid release following glucose ingestion, we tested the hypothesis that bile acids influence peripheral glucose uptake. We could not detect an effect of bile acids on basal or insulin stimulated glucose uptake, however, in cultured adipocytes (data not shown). At present the physiologic role of sustained bile acid release following glucose ingestion is unknown.

Although insulin resistance is traditionally defined as the reduced ability of insulin to promote glucose uptake or as the impaired suppression of gluconeogenesis, resistance can emerge in other insulin-dependent processes. In prior studies, for example, inadequate suppression of lipolysis was observed in women with a history of gestational diabetes (Chan et al., 1992), and an elevated proteolysis rate was seen in individuals with HIV-associated insulin resistance (Reeds et al., 2006). In obesity, manifestations of insulin resistance include elevated rates of lipolysis (Robertson et al., 1991) and proteolysis (Jensen and Haymond, 1991; Luzi et al., 1996). We have used metabolic profiling to monitor

insulin action across multiple axes (Figure 4A), and with markers of each axis in hand, were able to detect kinetic differences between them. In healthy individuals, insulin's suppression of lipolysis and ketogenesis is rapid compared to its suppression of proteolysis (Figure 4C), consistent with the different concentrations of insulin required to inhibit each of these processes (Fukagawa et al., 1985; Nurjhan et al., 1986). Such differences in the activation thresholds of metabolic pathways could be contributing to the diversity in the clinical presentation of insulin resistance. Mouse models of insulin-resistant T2DM provide another example for inter-pathway differences in insulin sensitivity: in the livers of these mice, gluconeogenesis is resistant to suppression by insulin, while lipogenesis is responsive, resulting in hyperglycemia and hypertriglyceridemia (Brown and Goldstein, 2008). Collectively, these observations demonstrate that environmental or genetic perturbations could lead to selectivity in insulin sensitivity and contribute to pathogenesis.

Our study demonstrates that an individual's "insulin response profile", namely, the vector of sensitivities to insulin action along multiple physiologic axes, can be revealed by metabolite excursions in response to a glucose challenge. Specifically, we have shown that excursions in Leu/Ile and glycerol, reflecting the sensitivity of proteolysis and lipolysis to the action of insulin, are jointly predictive of fasting insulin, with each of the two excursions offering complementary and significant explanatory power (Figure 5B). These findings demonstrate that in two individuals, different axes could be responsible for the same elevation of fasting insulin, and simultaneous measurement of metabolites can distinguish between the two conditions. Monitoring of the response to glucose ingestion across multiple axes in larger, prospective clinical studies of pre-diabetics could establish links between insulin sensitivity profiles and disease progression, thus helping to predict future diabetes and its complications as well as to guide therapeutic interventions.

Methods

Human Subjects.

MACS (Metabolic Abnormalitites in College Studnets). Subjects were young adults in the age range 18-30 who volunteered for the study during the academic year 2006-7. Control subjects (ingesting water instead of glucose solution) were assigned at random, balancing gender. Metabolic profiling analysis was limited to those subjects with normal fasting glucose concentrations (below 100 mg/dL) and normal glucose tolerance (2-hr glucose concentration below 140 mg/dL).

FOS (Framingham Offspring Study). Subjects were selected at random (balancing gender) from among all participants in the fifth FOS examination cycle (1991-1995) aged 40-49 who had no diabetes mellitus, hypertension or prior cardiovascular disease. Additional selection criteria for FOS-NGT were normal fasting glucose concentrations and normal glucose tolerance. Additional selection criterion for FOS-IGT was impaired glucose tolerance (2-hr glucose concentration between 140 and 199 mg/dL).

Table 1 lists demographic and metabolic characteristics of groups from the MACS and FOS studies.

Oral Glucose Tolerance Test.

MACS. Subjects were admitted to the Massachusetts Institute of Technology Clinical Research Center (CRC) after a 10 hour overnight fast. An intravenous catheter was inserted into an antecubital vein or a wrist vein and fasting samples were drawn. Next, each subject ingested a glucose solution (Trutol, 75 g in 296 ml; NERL Diagnostics, East Providence, RI; contains citric acid and sodium benzoate) or an identical volume of bottled water (Poland

Spring Water, Wilkes Barre, PA) over a 5 minute period. Additional blood samples were drawn from the inserted catheter 30, 60, 90 and 120 minutes after ingestion. Subjects remained at rest throughout the test. The study protocol was approved by the MIT Committee on the Use of Humans as Experimental Subjects and the CRC Scientific Advisory Committee.

FOS. Subjects underwent an OGTT as part of the Framingham Offspring Study (Arnlov et al., 2005). After 12 hour overnight fast, subjects ingested 75g glucose in solution. Blood samples were drawn fasting and 120 minutes after glucose ingestion. The study protocol was approved by the Institutional Review Board at Boston Medical Center.

Metabolic Profiling.

Blood processing. Blood was drawn into EDTA coated tubes. In MACS, blood was centrifuged for 10 minutes at 6°C and 2,000 g. In FOS, blood was centrifuged for 30 minutes at 4°C and 1,950 g. Plasma samples were stored at -80°C.

Sample preparation and analysis. Plasma samples were thawed gradually, and 165 µL from each sample was mixed with 250 µL of ethanol solution (80% ethanol, 19.9% H_2O, 0.1% formic acid) to precipitate out proteins. After 2 hours at 4°C, the samples were centrifuged at 15,000 g for 15 minutes, and 300 µL of the supernatant was extracted and evaporated under nitrogen. Samples were reconstituted in 60 µL HPLC-grade water, and separated on three different HPLC columns. The columns were connected in parallel to a triple quadrupole mass spectrometer (4000 Q Trap, Applied Biosystems) operated in selected reaction monitoring mode. Each metabolite was identified by a combination of chromatographic retention time, precursor ion mass and product ion mass. Metabolite quantification was performed by integrating the peak areas

of product ions using MultiQuant software (Applied Biosystem). Additional details on the analytical methodology are provided in Supplementary data 1.

Glucose and Insulin.

Plasma glucose concentration was measured with a hexokinase assay (MACS: Quest Diagnostics, Cambridge, MA. FOS: Abbott Laboratories, IL). Insulin international units were determined using a radioimmunoassay (Diagnostic Product Corporation, Los Angeles, CA). In MACS, sodium fluoride – potassium oxalate blood tubes were used for glucose analysis, and blood tubes with no additive were used for insulin analysis.

Statistical Analysis.

Univariate analysis. The significance of a change from the fasting metabolite level was calculated using the paired Wilcoxon signed-rank test. The significance of a difference between glucose and water ingestion was calculated using the unpaired Wilcoxon rank sum test. Where all ~100 detected metabolites were tested, a significance threshold of $\alpha=0.001$ was used to account for multiple hypotheses testing. A threshold of $\alpha=0.05$ was used elsewhere.

Multivariate Analysis and Linear Regression. Two multivariate approaches were used to identify combinations of 2-hr metabolite changes predictive of fasting insulin: forward stepwise linear regression and partial least squares (PLS) regression. In stepwise linear regression, the significance threshold for addition of a variable to the model was $\alpha=0.05$. In PLS regression, the number of components to include in the model was determined using leave-one-out cross-validation. The prediction error of a model was expressed as the root mean square error of prediction (RMSEP). Missing values (less than 2% of the data) were replaced with the mean of present values.

Statistical analysis was performed in R (The R Project for Statistical Computing) and in Matlab (The MathWorks, Inc).

Acknowledgements

We thank Joseph Avruch, Ronald Kahn, Barbara Kahn, Sudha Biddinger, and Mark Herman for helpful comments on the manuscript; Toshimori Kitami for for glucose uptake measurements; and Alice M. McKenney for assistance in preparing figures. OS was supported by a training grant for Bioinformatics and Integrative Genomics from the National Human Genome Research Institute. TJW was supported by the National Institutes of Health (NIH) (R01-HL-086875 and R01-HL-083197). GDL was supported by the Heart Failure Society of America and the Harvard/MIT Clinical Investigator Training Program. REG was supported by the NIH (U01HL083141), the Donald W. Reynolds Foundation, and Fondation Leducq. VKM was supported by a Burroughs Wellcome Career Award in the Biomedical Sciences and a Howard Hughes Medical Institute Early Career Physician Scientist Award. This work was supported by a generous grant from the Broad Institute Scientific Planning and Allocation of Resources Committee, a General Clinical Research Center grant awarded by the NIH to the Massachusetts Institute of Technology General Clinical Research Center (MO1-RR01066), and a NIH / National Heart Lung and Blood Institute contract supporting the Framingham Heart Study (N01-HC-25195).

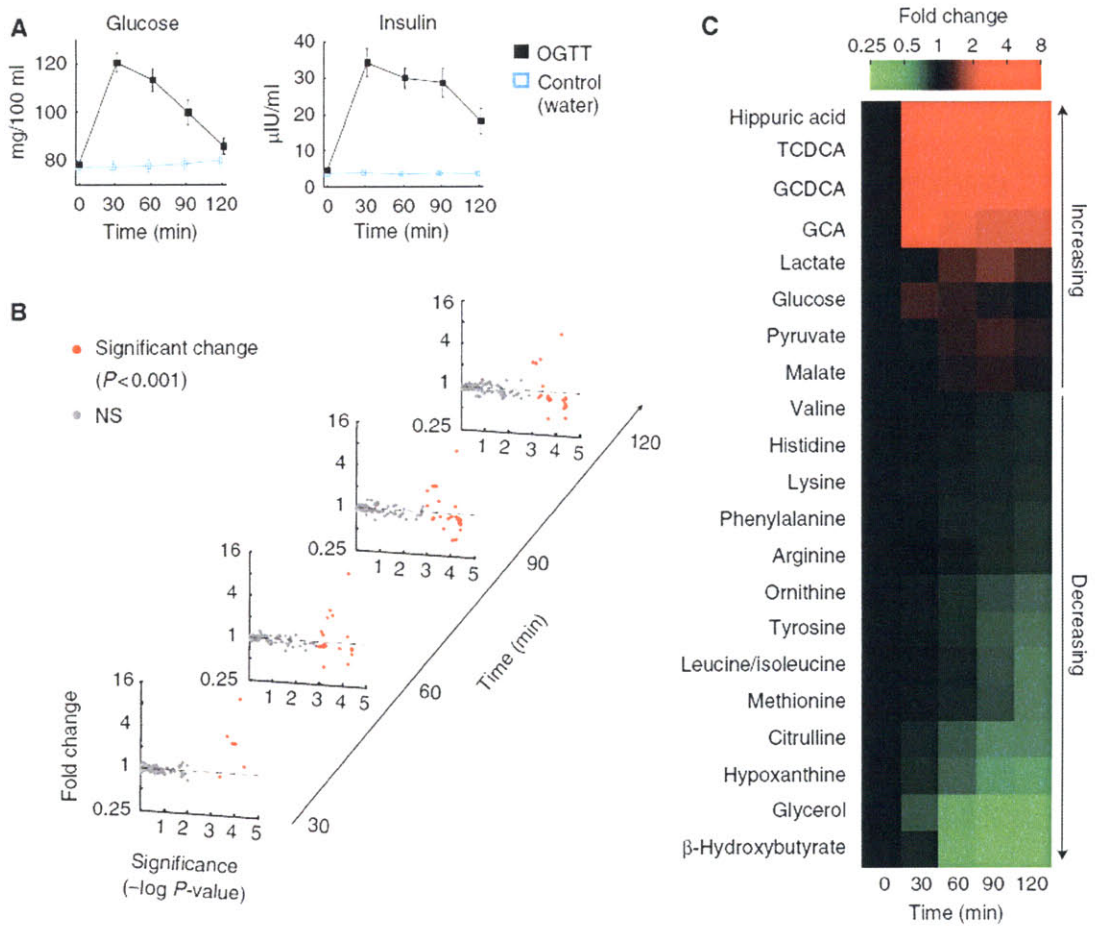


Figure 1: Temporal response to an oral glucose challenge in individuals with normal glucose tolerance (MACS).

(A) Kinetics of blood glucose and insulin in response to glucose ingestion (mean ± s.e.m.).

(B) Magnitude and significance of metabolite change over time. Dots represent the 97 metabolites detected in plasma. Change is with respect to the fasting metabolite levels. Significant ($p<0.001$) changes are colored red.

(C) Significant metabolite changes. 21 metabolites changed significantly ($p<0.001$) when compared to their fasting levels and showed a significantly ($p<0.05$) distinct response compared to control (water ingestion). Color intensity reflects the median fold change compared to the fasting levels. Metabolites are ordered according to the magnitude of change.

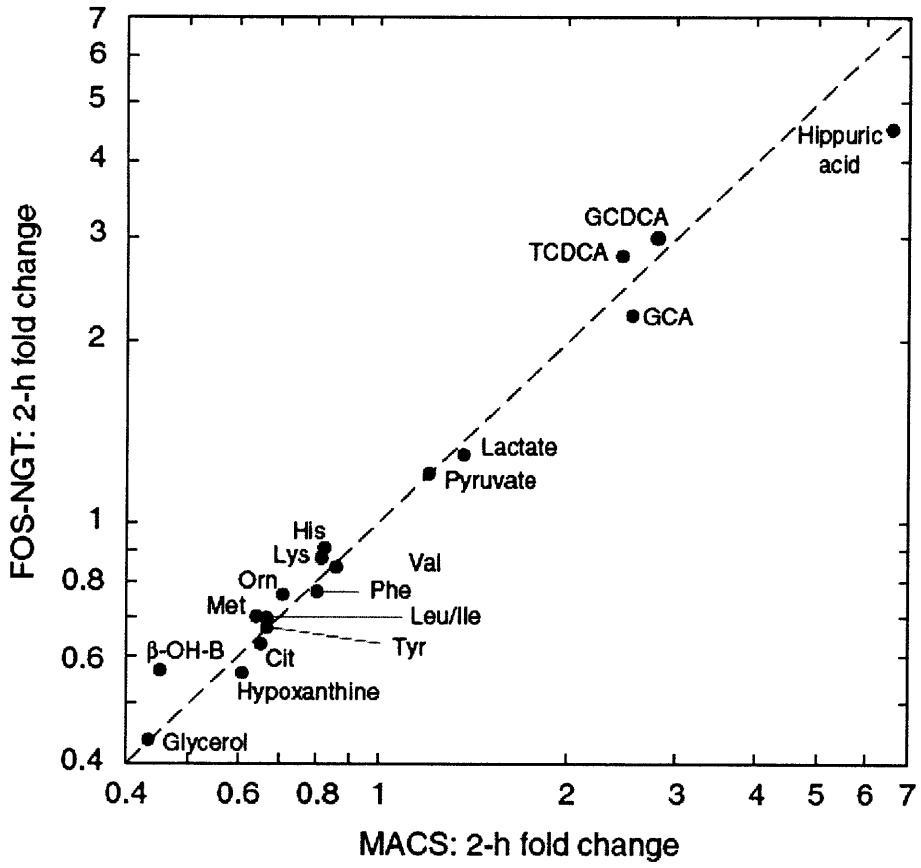


Figure 2: Validation of metabolite changes at the 2-hr time point.

18 out of the 20 metabolites that changed significantly in MACS (Figure 1C) replicated ($p<0.05$) in FOS-NGT. Dots correspond to the median fold change of metabolites at the 2-hr time point. Abbreviations: TCDCA, taurochenodeoxycholic acid; GCDCA, glycochenodeoxycholic acid; GCA, glycocholic acid, Orn: ornithine, Cit: citrulline, β-OH-B: β-hydroxybutyrate.

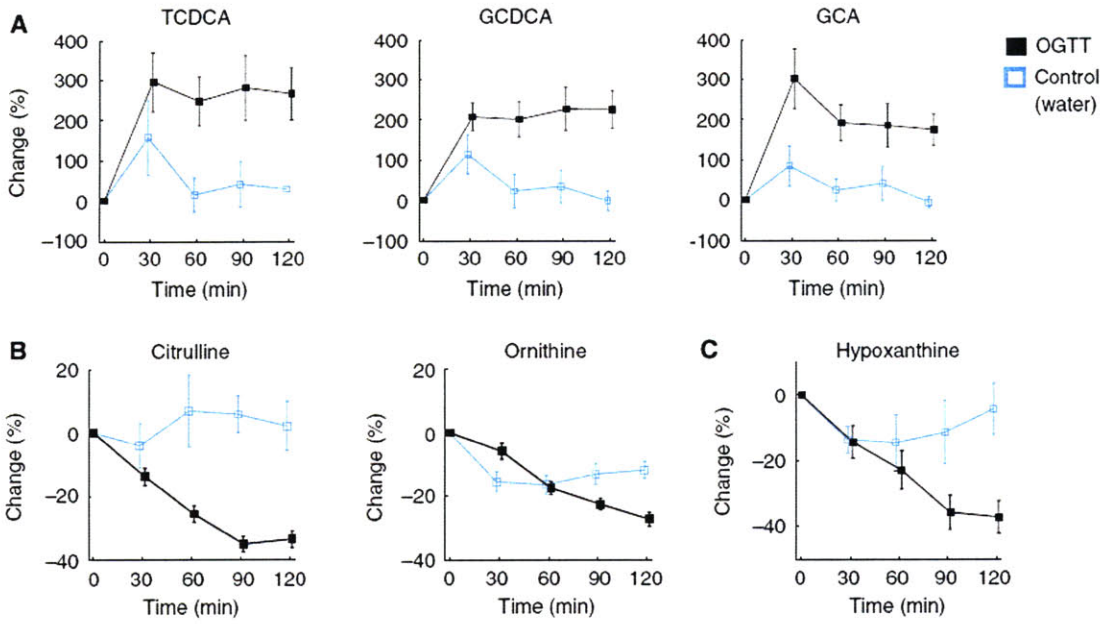


Figure 3: Metabolic responses not previously linked to glucose homeostasis.

Kinetic patterns in the MACS group are shown (mean ± s.e.m.).

(A) Bile acids. Abbreviations: TCDCA, Taurochenodeoxycholic acid; GCDCA, Glycochenodeoxycholic acid; GCA, Glycocholic acid.

(B) Citrulline and ornithine, urea cycle intermediates.

(C) Hypoxanthine, a product of purine nucleotide degradation.

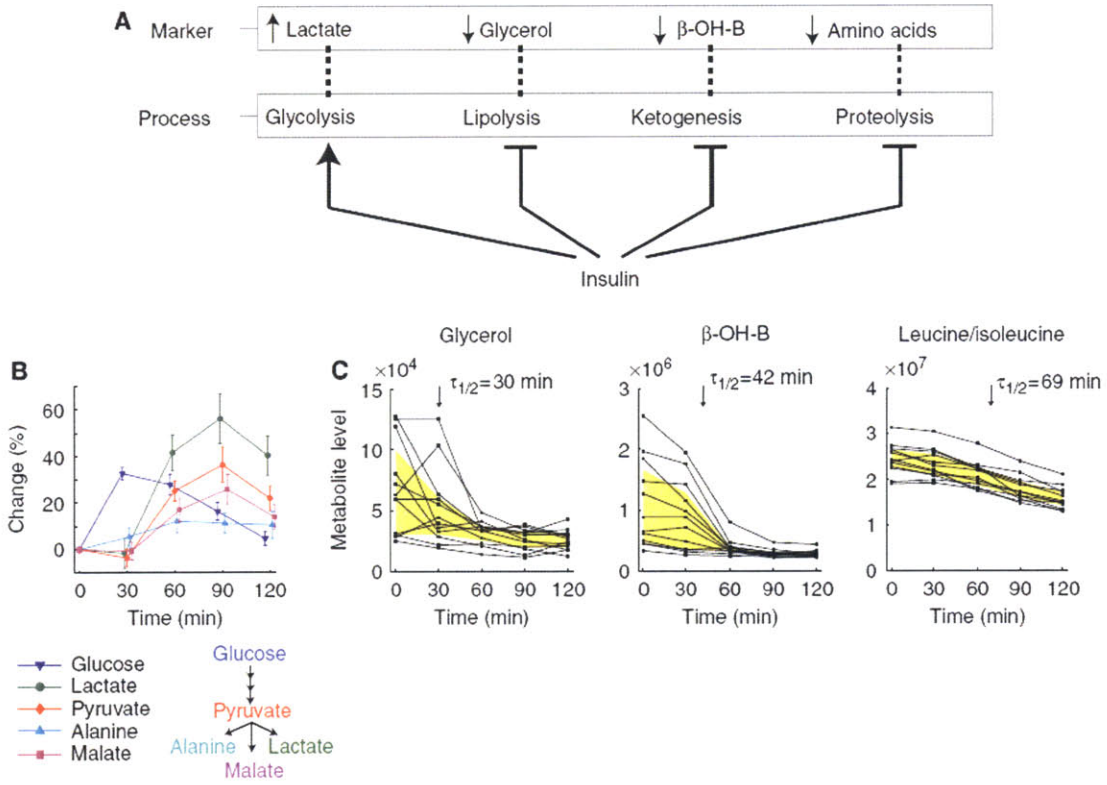


Figure 4: Metabolites reflecting four axes of insulin action.

(A) Four axes of insulin action and their associated metabolic markers.

(B) The kinetics of glucose and pyruvate derivatives (MACS, mean ± s.e.m.).

(C) The kinetics of insulin's suppression of catabolism. Each line corresponds to one of 12 individual MACS subjects. $\tau_{1/2}$: time to half-maximal decrease (median of all subjects). The inter-quartile range of metabolite levels is shaded in yellow. Abbreviations: β-OH-B, β-hydroxybutyrate.

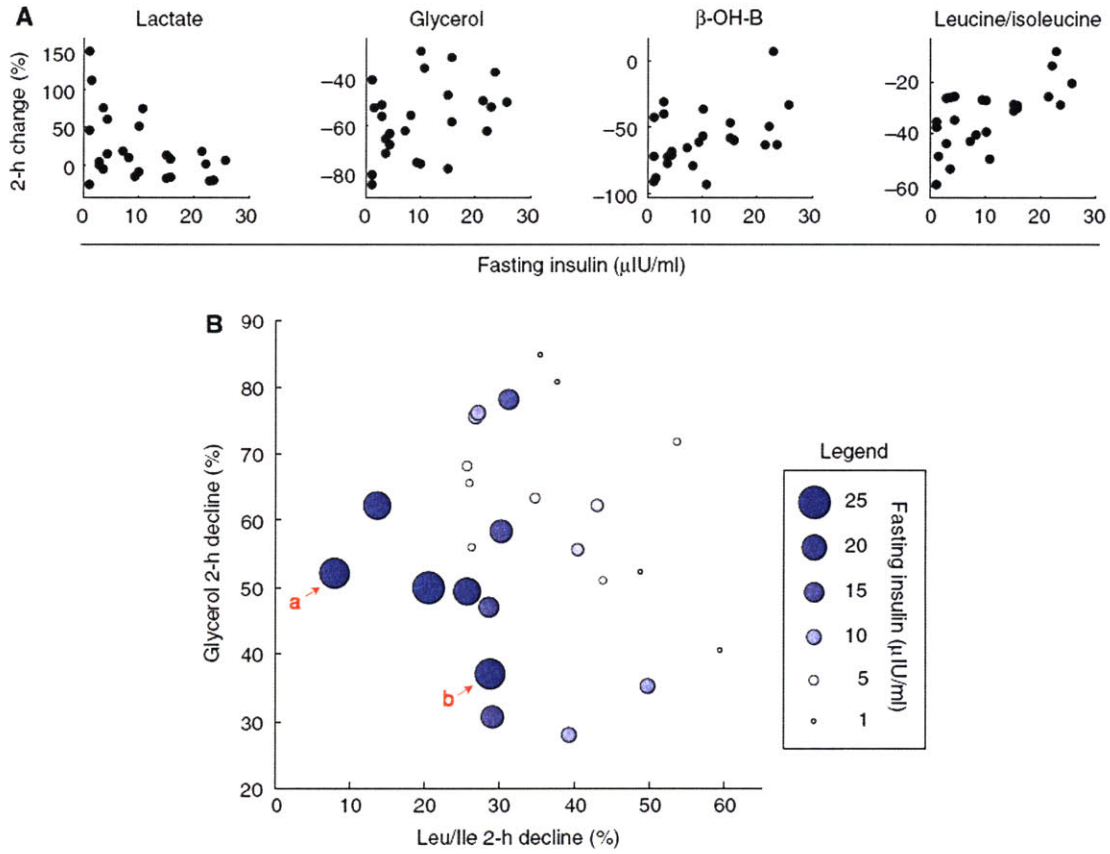


Figure 5: Correlation between fasting insulin and 2-hr metabolite changes in individuals with impaired glucose tolerance (FOS-IGT)

(A) 2-hr changes in markers of insulin action are correlated with fasting insulin concentration. Each dot corresponds to an individual.

(B) A bi-variate model explaining fasting insulin using the 2-hr decline of Leu/Ile and glycerol. Each circle represents an individual, and the circle size is proportionate to fasting insulin levels. [a]A representative individual exhibiting a blunted decline in Leu/Ile (resistant to proteolysis suppression). [b]A representative individual exhibiting a blunted decline in glycerol (resistant to lipolysis suppression).

Clinical Study:	MACS		FOS	
Group:	Glucose[a] (n=22)	Water[b] (n=7)	FOS-NGT (n=25)	FOS-IGT (n=25)
Age	23 ± 3 (18-30)	24 ± 4 (20-30)	45 ± 3 (40-49)	46 ± 3 (40-50)
Gender	9 ♀, 13 ♂	3 ♀, 4 ♂	13 ♀, 12 ♂	13 ♀, 12 ♂
Ancestry	Wh:9,As:6,Un:7	Wh:3,Aa:1,As:1,Un:2	Wh:25	Wh:25
BMI	22.4 ± 2.1 (18.3-26.9)	22.1 ± 2.7 (17.8-26.0)	24.6 ± 3.4 (19.0-31.5)	26.8 ± 4.8 (18.8-41.2)
Fasting Glucose (mg/dL)	78 ± 5 (71-90)	77 ± 7 (70-86)	89 ± 6 (76-100)	100 ± 9 (87-115)
120 min. Glucose (mg/dL)	86 ± 16 (66-119)	80 ± 9 (71-92)	88 ± 21 (43-122)	153 ± 12 (140-180)
Fasting Insulin (uIU/mL)	4.6 ± 2.9 (2.8-14.2)	3.6 ± 0.7 (2.8-4.8)	4.2 ± 2.7 (1.0-10.7)	10.3 ± 8.1 (1.0-25.7)
120 min. Insulin (uIU/mL)	18.1 ± 16.5 (3.0-75.9)	3.6 ± 1.0 (2.9-5.5)	29.7 ± 20.5 (1.0-93.3)	102.8 ± 51.8 (35.0-202.3)
IGT / NGT[c]	0 / 22	N/A	0 / 25	25 / 0

Quantitative variables are expressed as mean ± s.d. (range).

Abbreviations: MACS, Metabolic Abnormalities in College Students, conducted at MIT Clinical Research Center; FOS, Framingham Offspring Study; BMI, body mass index. Ancestry abbreviations: Wh, White; As, Asian; Aa, African American; Un, Unknown

[a] Subjects ingesting glucose (OGTT).

[b] Subjects ingesting water (control).

[c] Numbers of subjects in each glucose tolerance category. NGT, normal glucose tolerance; IGT, impaired glucose tolerance (American Diabetes Association, 2007).

Table I: Demographic and Clinical Characteristics of Human Subjects

Predictor(s)	R^2_{adj}	p-value	Prediction Error[a]
Δ[b] Leucine/Isoleucine	0.36	9E-4	6.65
Δ Valine	0.17	3E-2	7.74
Δ Lactate	0.16	3E-2	7.60
Δ Glycochenodeoxycholic acid	0.14	4E-2	7.86
Δ Methionine	0.14	4E-2	7.68
Δ β-hydroxybutyrate	0.14	4E-2	7.90
Δ Leucine/Isoleucine + Δ Glycerol[c]	0.54	7E-5	5.66
PLS[d]	0.46	1E-4	6.89
BMI	0.33	1E-3	6.74

[a] The prediction error is expressed as the root of mean square error of prediction (RMSEP), in μIU/mL insulin.

[b] Δ denotes log of the 2-hr fold change of metabolite levels.

[c] A bivariate model consisting of the 2-hr changes in leucine/isoleucine and in glycerol

[d] Partial least squares based on changes in the 18 validated metabolites.

Table II: Regression Models Relating Fasting Insulin to 2-hr Metabolite Change in Individuals with Impaired Glucose Tolerance (FOS-IGT)

References

American Diabetes Association 2007. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 30(Supplement 1):S42-S47.

Angelin B, Bjorkhem I, Einarsson K, Ewerth S. 1982. Hepatic uptake of bile acids in man. Fasting and postprandial concentrations of individual bile acids in portal venous and systemic blood serum. J Clin Invest 70(4):724-731.

Arnlov J, Pencina MJ, Nam BH, Meigs JB, Fox CS, Levy D, D'Agostino RB, Vasan RS. 2005. Relations of insulin sensitivity to longitudinal blood pressure tracking: variations with baseline age, body mass index, and blood pressure. Circulation 112(12):1719-1727.

Brauer MJ, Yuan J, Bennett BD, Lu W, Kimball E, Botstein D, Rabinowitz JD. 2006. Conservation of the metabolomic response to starvation across two divergent microbes. Proc Natl Acad Sci U S A 103(51):19302-19307.

Brown MS, Goldstein JL. 2008. Selective versus Total Insulin Resistance: A Pathogenic Paradox. Cell Metab 7(2):95-96.

Cahill GF, Jr. 2006. Fuel metabolism in starvation. Annu Rev Nutr 26:1-22.

Chan SP, Gelding SV, McManus RJ, Nicholls JS, Anyaoku V, Niththyananthan R, Johnston DG, Dornhorst A. 1992. Abnormalities of intermediary metabolism following a gestational diabetic pregnancy. Clin Endocrinol (Oxf) 36(4):417-420.

De Barros SG, Balistreri WF, Soloway RD, Weiss SG, Miller PC, Soper K. 1982. Response of total and individual serum bile acids to endogenous and exogenous bile acid input to the enterohepatic circulation. Gastroenterology 82(4):647-652.

Devlin TM. 2002. Textbook of Biochemistry with Clinical Correlations. 5th ed. New York: Wiley-Liss.

Dunn WB, Bailey NJ, Johnson HE. 2005. Measuring the metabolome: current analytical technologies. Analyst 130(5):606-625.

Felig P. 1975. Amino acid metabolism in man. Annu Rev Biochem 44:933-955.

Fukagawa NK, Minaker KL, Rowe JW, Goodman MN, Matthews DE, Bier DM, Young VR. 1985. Insulin-mediated reduction of whole body protein breakdown. Dose-response effects on leucine metabolism in postabsorptive men. J Clin Invest 76(6):2306-2311.

Hanson RL, Pratley RE, Bogardus C, Narayan KM, Roumain JM, Imperatore G, Fagot-Campagna A, Pettitt DJ, Bennett PH, Knowler WC. 2000. Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. Am J Epidemiol 151(2):190-198.

Jensen MD, Haymond MW. 1991. Protein metabolism in obesity: effects of body fat distribution and hyperinsulinemia on leucine turnover. Am J Clin Nutr 53(1):172-176.

Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. 1979. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 110(3):281-290.

Kelley D, Mitrakou A, Marsh H, Schwenk F, Benn J, Sonnenberg G, Arcangeli M, Aoki T, Sorensen J, Berger M, Sonksen P, Gerich J (1988) Skeletal muscle glycolysis, oxidation, and storage of an oral glucose load. J Clin Invest 81:1563-1571

Kubota K, Ishizaki T. 1991. Dose-dependent pharmacokinetics of benzoic acid following oral administration of sodium benzoate to humans. Eur J Clin Pharmacol 41(4):363-368.

Liddle RA, Goldfine ID, Rosen MS, Taplitz RA, Williams JA. 1985. Cholecystokinin bioactivity in human plasma. Molecular forms, responses to feeding, and relationship to gallbladder contraction. J Clin Invest 75(4):1144-1152.

Luzi L, Castellino P, DeFronzo RA. 1996. Insulin and hyperaminoacidemia regulate by a different mechanism leucine turnover and oxidation in obesity. Am J Physiol 270(2 Pt 1):E273-281.

Mateos FA, Puig JG, Jimenez ML, Fox IH (1987) Hereditary xanthinuria. Evidence for enhanced hypoxanthine salvage. J Clin Invest 79: 847-852

Nurjhan N, Campbell PJ, Kennedy FP, Miles JM, Gerich JE. 1986. Insulin dose-response characteristics for suppression of glycerol release and conversion to glucose in humans. Diabetes 35(12):1326-1331.

Reeds DN, Cade WT, Patterson BW, Powderly WG, Klein S, Yarasheski KE. 2006. Whole-body proteolysis rate is elevated in HIV-associated insulin resistance. Diabetes 55(10):2849-2855.

Robertson DA, Singh BM, Nattrass M. 1991. Effect of obesity on circulating intermediary metabolite concentrations in the absence of impaired glucose tolerance. Int J Obes 15(10):635-645.

Sabatine MS, Liu E, Morrow DA, Heller E, McCarroll R, Wiegand R, Berriz GF, Roth FP, Gerszten RE. 2005. Metabolomic identification of novel biomarkers of myocardial ischemia. Circulation 112(25):3868-3875.

Watanabe M, Houten SM, Mataki C, Christoffolete MA, Kim BW, Sato H, Messaddeq N, Harney JW, Ezaki O, Kodama T, Schoonjans K, Bianco AC, Auwerx J. 2006. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. Nature 439(7075):484-489.

Yamaoka T, Kondo M, Honda S, Iwahana H, Moritani M, Ii S, Yoshimoto K, Itakura M (1997) Amidophosphoribosyltransferase limits the rate of cell growth-linked de novo purine biosynthesis in the presence of constant capacity of salvage purine biosynthesis. J Biol Chem 272: 17719-17725

Chapter 4

—

A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent cell culture media

Oded Shaham, Olga Goldberger, Arvind Ramanathan, Clary B. Clish, Katherine B. Sims and Vamsi K. Mootha

Corresponding Supplementary Material can be found in Appendix B.

A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent cell culture media

Mitochondrial respiratory chain disorders (RCD) are the most common inborn error of metabolism, but their pathogenesis is poorly understood and they are notoriously difficult to diagnose owing to their highly variable clinical presentation and genetic heterogeneity. To rigorously define a biochemical signature of RCD, we applied mass spectrometry-based metabolic profiling to both a cellular model of mitochondrial dysfunction and to the plasma of RCD patients. In cultured myotubes, we found that the secretion or consumption patterns of 32 metabolites were altered by inhibition of the mitochondrial respiratory chain. We hypothesized that this cell culture media-defined panel of metabolites may reflect the biochemical derangements of plasma from RCD patients. Indeed we found that the media-defined panel of metabolites strongly discriminated between plasma of RCD patients (n=17) and healthy individuals (n=25), far better than random sets of metabolites ($P<0.01$). Nine of these metabolites accounted for the bulk of this signal, were concordantly altered in human plasma and in culture media, and included markers of impaired substrate oxidation (elevated lactate, alanine, 2-oxoglutarate, succinate, fumarate, malate), blunted pyrimidine biosynthesis (decreased uridine), reduced energy charge (elevated creatine), and increased choline. Importantly, a linear combination of these 9 metabolites discriminated between RCD patients and healthy controls in a manner superior to and complementary to currently used plasma markers. Our study provides proof of concept that profiling spent media from cellular models of disease may provide a window into the metabolic derangements in plasma, an approach that could in principle be extended to complex human diseases.

Introduction

The respiratory chain (RC) of mammalian cells comprises a series of five enzymatic complexes embedded in the inner mitochondrial membrane that perform oxidative phosphorylation to regenerate adenine triphosphate (ATP). Respiratory chain activity affects cellular physiology in multiple ways, including physical coupling to biochemical reactions and control of key regulatory parameters such as the energy charge and the [NADH]/[NAD^+] ratio. In humans, defects in mitochondrial or nuclear genes that encode respiratory chain proteins, or factors necessary for its maturation and assembly, lead to respiratory chain disease (RCD). It is estimated that RCD affects at least 18.4 in 100,000 people (1) and is the most common group of inborn errors of metabolism (2).

The diagnosis of RCD is challenging owing to its genetic heterogeneity and variable clinical presentation. A number of diagnostic algorithms have been proposed (3-5) that integrate clinical, biochemical, histological and molecular findings. Abnormal concentrations of several metabolites in biological fluids represent a minimally-invasive component of these algorithms, including the elevation of lactate, alanine, pyruvate and the lactate/pyruvate ratio. TCA cycle intermediates have also been proposed as indicative of RCD (6, 7), as well as various other organic acids, amino acids and acylcarnitines (7, 8), though these suggestions are based on isolated case reports and have not been incorporated into diagnostic algorithms. Little information is available on the diagnostic accuracy of metabolite markers or on the correlation amongst them. Lactate, the most widely used metabolic marker, unfortunately is often not elevated in RCD patients, and may be elevated due to various other causes unrelated to RCD (8, 9). A systematic evaluation of metabolic alterations in respiratory chain dysfunction is needed in order to identify useful diagnostic markers - such markers could lead to increased accuracy of RCD diagnosis, and may also reveal the cellular pathways contingent on RC activity.

The extracellular metabolic profile in cell culture provides information on consumption and secretion of metabolites, and has been previously used as an innovative method for rapidly classifying yeast mutants that were otherwise indistinguishable from wild type (10). Inspired by this approach, we hypothesized that the culture media profile of a cellular RCD model could similarly capture features of the metabolic derangements in the plasma of RCD patients. This hypothesis stems from the notion that the net sum of secretion and consumption by all body cells determines the concentration of metabolites in plasma. We have recently developed a targeted mass spectrometry method capable of measuring 191 metabolites in biological samples (11, 12). Importantly, this method spans a wide biochemical spectrum including amino acids and nitrogenous compounds, organic acids, nucleotides and sugars, enabling simultaneous monitoring of multiple pathways. Here, we apply this technology to derive an extracellular profile of RC dysfunction from a cellular model of mitochondrial dysfunction, and then demonstrate that this panel of metabolites can be utilized to define a plasma biochemical signature of human RCD.

Results

A Metabolic Signature of RC Dysfunction in a Myotube Culture Model

We induced RC dysfunction in differentiated myotubes using rotenone and antimycin, potent inhibitors of NADH dehydrogenase (complex I) and ubiquinol:cytochrome C oxidoreductase (complex III), respectively. We chose myotubes as a model system due to their high oxygen requirement, and since myopathy is a frequent manifestation of RCD. The inhibitor concentrations we used were effective, as apparent from the decrease in the rate of oxygen consumption (Figure 1A), but did not decrease cell viability even after 24 hours of drug treatment. We selected the incubation time (8 hours), cell number (5×10^6) and media volume (10 ml) so as to produce at least a two-fold elevation of lactate levels in the media compared to the vehicle control (DMSO). At the

concentrations used, rotenone and antimycin elicited comparable effects on respiratory chain activity, as demonstrated by a similar decrease in oxygen consumption (Figure 1A) and a similar increase in lactate secretion (Figure 1B).

In the absence of any drug treatment, the levels of 25 metabolites significantly ($p<0.05$) changed from fresh to spent media, reflecting consumption and secretion by cells. We detected the consumption of 8 metabolites, 6 of which are nutrients provided in the media, including amino acids (serine, tryptophan, tyrosine, phenylalanine and leucine/isoleucine) and choline. The consumption of glucose, glutamine and other amino acids was below the significance threshold, likely due to their high starting concentration in media. We detected the secretion of 17 metabolites, including amino acids that can be synthesized by mammalian cells (alanine, asparagine, ornithine, proline, aspartate and glutamate), nucleosides (cytidine and uridine), tricarboxylic acid (TCA) cycle intermediates (malate, 2-oxoglutarate, aconitate, succinate and citrate), lactate and creatine.

We next determined which metabolites uptake or release patterns were significantly influenced by the respiratory chain inhibitors. We found that RC inhibition altered significantly ($p<0.05$) the levels of 32 metabolites (Figure 1B). Cultures treated with RC inhibitors exhibited faster glucose consumption and lactate secretion, reflecting acceleration of non-oxidative glucose metabolism consistent with a dysfunction of mitochondrial ATP synthesis. In contrast, the uptake of amino acids and choline decreased, suggesting slower biosynthesis of proteins and membrane lipids, respectively.

We found multiple differences between the antimycin and the rotenone treatments (Figure 1B). Two metabolites, succinate and uridine, exhibited a particularly striking discordance between the two inhibitors, being significantly altered by both but in opposite directions (Figure 1B). Succinate secretion was strongly enhanced with antimycin but blunted with rotenone. The succinate pattern, as well as the higher secretion of fumarate and malate with rotenone

compared to antimycin, suggest an opposite effect of the two inhibitors on succinate dehydrogenase (complex II) activity: whereas antimycin blocks all enzymes transferring electrons to ubiquinone by preventing its re-oxidation, rotenone disables only complex I (Figure 1C). It has been previously observed that complex I inhibition enhances complex II activity, presumably by making a larger part of the coenzyme Q pool available (13). We can thus explain the patterns of succinate, fumarate and malate by modulation of complex II activity. Dihydroorotate dehydrogenase (DHODH), a key enzyme in pyrimidine biosynthesis, interacts with coenzyme Q similarly to complex II (14, 15), and is therefore expected to be affected by antimycin and rotenone in the same fashion. Indeed, the secretion of uridine, a product of this pathway, was elevated with rotenone, possibly reflecting higher activity of DHODH, and decreased with antimycin, suggesting slower de-novo pyrimidine synthesis (Figure 1C). In addition, higher dependence on uridine salvage for pyrimidine nucleotide synthesis could contribute to the lower uridine secretion with antimycin. In accordance with the media profile, intracellular levels of pyrimidine nucleotides, but not purine nucleotides, were significantly decreased in the myotube cultures treated with antimycin (Figure S1). Together, succinate and uridine provide a metabolic distinction between a "complex I only" dysfunction and an "all respiratory chain" dysfunction.

Cell Culture-Derived Metabolites Distinguish RCD Patients from Healthy Controls

To identify biochemical alterations in human RCD, we performed metabolic profiling on plasma from 16 patients (Table 1) and 25 healthy controls. All RCD patients had a pathogenic mutation (7 patients) and/or an abnormally low RC enzyme activity in muscle (11 patients). The cohort included 4 patients with Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke (MELAS, mt A3243G), 1 patient with Myoclonus, Epilepsy and Ragged-Red Fibers (MERRF, mt A8344G), 2 patients with mitochondrial DNA (mtDNA) deletions, and one

patient with mtDNA depletion. Muscle involvement was present in all patients, either as part of the clinical presentation or as a biochemical feature (Table 1). Plasma was obtained during a routine outpatient visit. We detected 87 metabolites in plasma, including 29 of the 32 metabolites altered in culture media. Cytidine, inosine and pentose-phosphate were not detected in plasma. We next asked whether the 29 cell culture-derived metabolites detected in plasma could distinguish the RCD patients from the healthy controls. Unsupervised principal component analysis in the space of the 29 model-derived metabolites showed excellent separation of patients from controls (Figure 2A). Interestingly, two plasma samples from the same patient (labeled 6.1 and 6.2, Figure 2A) obtained on two separate visits six months apart were located close together in the space of the first two principal components, indicating stability of the metabolic profile. To test whether the separation between patients and controls was unique to the model-derived metabolite set, we evaluated 1000 other metabolite sets of the same size selected randomly from all metabolites detected in human plasma. We calculated the discriminatory power of each metabolite set using linear classification and cross-validation, taking the area under the receiver operating characteristic (ROC) curve as a measure of performance. While the average performance with random metabolite sets was 0.87, the model-derived set achieved a performance of 0.98, higher than 99.1% of all random sets (Figure 2B). The model-derived set ranked similarly high when performance was expressed as the sensitivity or the specificity. These findings suggest that the extracellular metabolic profile of RC inhibition in cultured cells reflects the plasma biochemical profile of RCD patients.

Plasma Markers of RCD

Next we asked which individual metabolite trends are shared between media of RC-inhibited myotubes and plasma of RCD patients. Nine metabolites exhibited a significant ($p<0.05$) difference between RCD patients and controls, changed significantly in response to myotube inhibition, and showed the same

trend (elevation or decrease) in plasma and in media (Figure 3A). The weights of these 9 metabolites were among the highest in the linear classifier based on all 29 culture-derived metabolites (Table S1), and the performance (area under ROC curve) of a classifier based on the 9 metabolites was 0.96, only slightly lower than the 29-metabolites classifier (Figure 2B). Thus these 9 metabolites capture a substantial part of the discriminatory power of the culture-derived signature. The metabolites that were concordant between the plasma and media included the expected elevation of lactate and alanine, both of which are part of current diagnostic algorithms for RCD. This group also included 4 elevated TCA cycle intermediates: 2-oxoglutarate, succinate, fumarate and malate. The only metabolite that was decreased in the RCD patients was uridine, whose median level in RCD patients was 24% lower than the median of controls. Creatine showed the most striking elevation in RCD patients. To avoid the confounding effect of oral creatine supplementation, we excluded from the analysis all 5 patients whose medication list at the time of the test included creatine. Still, median creatine level in RCD patients was 233% higher than the median level in controls. Finally, choline was 18% elevated in RCD patients. Adjusting for gender, all 9 metabolites were significantly ($p<0.05$) different between patients and controls. We previously reported 21 plasma metabolites that change in response to glucose ingestion – the only overlapping metabolites are lactate and malate while the remaining 7 metabolites were measured but not influenced by a glucose challenge (12).

To assess whether these markers complement each other in discriminating between RCD patients and controls, we evaluated linear classifiers based on 3 different metabolite sets: set #1 consisted of lactate and alanine, the only two metabolites identified here which are currently incorporated into diagnostic algorithms (3-5); set #2 contained set #1 as well as the 4 TCA cycle intermediates identified; and set #3 included all 9 metabolites (Table 2). Set #2 achieved higher classification performance than set #1, indicating that TCA cycle intermediates carry useful information independent of lactate and alanine.

Performance increased further from set #2 to set #3, indicating that the remaining 3 metabolites, uridine, choline and creatine, contain additional independent information. The same ranking of the 3 sets was observed regardless of the performance statistic used (area under the ROC curve, sensitivity or specificity).

Discussion

In the current study we have derived an extracellular metabolic profile of respiratory chain dysfunction from a simple cell culture model of RCD. The use of a cellular model facilitated a well-powered and well-controlled search for the metabolic consequences of respiratory chain dysfunction. This enabled us to focus the subsequent clinical investigation on a smaller number of metabolites (Figure 1B). The concordance of the 9 plasma markers we identified here (Figure 3) with the trend in culture media provides strong evidence for a mechanistic link to respiratory chain dysfunction. Our study design was particularly useful given the limitations of clinical investigations of rare disorders – namely, the relatively small number of patients in the cohort and the challenges of potential confounders such as medication usage, diet, and age. Cell culture media obviously cannot model the complex regulation of metabolite concentrations in human plasma. Our finding that the cell culture media profile of RC inhibition recapitulates differences in plasma between patients and controls (Figure 2) suggests that consumption and secretion patterns in cell culture can capture, at least in part, the altered metabolic state in RCD patients. This approach could also be applicable to other metabolic diseases that can be adequately modeled in cell culture.

At present, the most widely used plasma markers of RCD are lactate, alanine and pyruvate, all of which reflect impaired pyruvate oxidation. Our results confirm the expected elevations in plasma lactate and alanine (pyruvate was not measured), but additionally reveal metabolite changes corresponding to other pathways that are contingent on the mitochondrial respiratory chain (Figure 3).

The elevation of TCA cycle intermediates in RCD patients likely reflects the diminished capacity of the RC to oxidize substrates originating from various pathways that feed into the TCA cycle. While elevated TCA cycle intermediates have been previously reported in RCD patients, their diagnostic utility when measured in conjunction with pyruvate oxidation markers has not been previously evaluated. Our results suggest that these metabolites contain discriminatory information complementary to lactate and alanine (Table 2). We have also identified 3 metabolite alterations not previously reported in RCD patients: high creatine and choline levels, and low uridine. The cellular balance between creatine and phosphocreatine reflects the capacity to regenerate ATP, and diminished phosphocreatine concentrations have been detected in the muscle of patients with mitochondrial myopathy (16). The elevated plasma creatine detected here could be an extracellular manifestation of the low energy charge in muscle. We speculate that the elevation of choline could reflect alterations in phospholipids metabolism.

Uridine is unique among the plasma RCD markers identified here in that it is the only one decreased (Figure 3). Other plasma metabolites previously reported to be decreased in RCD patients include citrulline (19) and free carnitine (20), and though we measured these metabolites in the current study we did not detect such decreases (N.B. We excluded from analysis patients taking a carnitine supplement). To our knowledge this is the first report of uridine deficiency in the plasma of RCD patients. It is notable that cells devoid of mtDNA require uridine supplementation for survival (17), presumably due to DHODH dysfunction (Figure 1C). In analogy, the low plasma uridine levels in RCD patients could indicate higher reliance on salvage of pyrimidines for maintaining cellular pyrimidine nucleotide levels, as well as less de-novo pyrimidine synthesis. While in cultured myotubes complex III inhibition diminished the intracellular pyrimidine nucleotide pools (Figure S1), it is unknown whether the same occurs *in vivo*. Uridine supplementation is beneficial in patients with a primary defect in the pyrimidine de-novo biosynthesis pathway (18). It is possible

that uridine supplementation could also be beneficial for patients with a uridine deficiency secondary to RCD.

Genetic lesions of the respiratory chain can affect the entire chain, or a specific enzymatic complex. Mutations in components of the mitochondrial translational machinery or in mtDNA maintenance factors affect the entire chain, while mutations in structural subunits or assembly factors of a particular complex affect one complex primarily. While histology or biochemical analysis of muscle biopsy can suggest which complexes are affected, there are currently no circulating biochemical markers that provide this information. Here, we describe some striking differences between the complex I and complex III media profiles. Some of these differences, as in the TCA cycle and pyrimidine biosynthesis pathways, can be explained at the levels of coenzyme Q and the dehydrogenases interacting with it (Figure 1C), while for other pathways the mechanism is presently unknown. Most of the RCD patients studied here had either a defect in complexes III or IV detected in muscle, or a mutation affecting the translation of mtDNA-encoded subunits in these complexes. Judging from the patterns of succinate and uridine, which exhibited a discordance between complex I and complex III inhibition (Figure 1B), the RCD patients as a group displayed an "all respiratory chain" dysfunction, with elevated succinate and decreased uridine (Figure 3), rather than a "complex I only" profile. Metabolic profiling of genetic RCD models, an example of which has been recently reported in *C. elegans* (21), could help delineate the pathogenesis of specific mutations. Future well-powered studies of human cohorts or genetic models may reveal genotype-specific metabolic phenotypes, and in particular contrast the "complex I only" profile with the "all respiratory chain" profile.

In summary, we have reported a plasma biochemical signature of human RCD, identified through systematic metabolic profiling of a cellular disease model. The RCD markers identified here reflect multiple pathways contingent on functional mitochondria, spanning substrate oxidation, biosynthesis and energy

charge. Importantly, we have found that these markers are not redundant but rather complementary to each other. While our results need to be validated in an independent study, controlling for potential confounding factors such as medications, diet and age, the combination of plasma metabolites identified here could improve the accuracy of minimally-invasive diagnosis of mitochondrial disease and provide a means for monitoring its progression. Beyond its value for detecting and monitoring disease, the signature of mitochondrial RCD we report here provides insight into the pathogenesis of this heterogenous group of human disorders. Importantly, our report provides a proof of concept that metabolic profiling of spent media from a cellular model of disease can provide a window into the metabolic derangements in the plasma of human patients. This approach may enjoy application to other, even complex metabolic disorders and could help understand their biochemical pathogenesis.

Materials and Methods

Respiratory Chain Inhibition in Cell Culture

Mouse C2C12 myoblasts were obtained from ATCC and grown in DMEM (Sigma) containing 4.5g/mL glucose with 10% FBS in 10 cm plates. To induce myotube differentiation, FBS was replaced with 2% horse serum (Sigma) for 4 days. On the day of the experiment, cultures were placed in 10 mL fresh media in the presence of 0.1 μM Rotenone (Sigma R8875), 0.5 μM antimycin (Sigma A8674) or DMSO, and after 8 hours of incubation spent media was sampled for metabolic profiling. The experiment was done in 8 biological replicates, and 8 samples of fresh media were also obtained for comparison to the spent media. To prepare media samples for LC-MS/MS analysis, 100 μL of media was mixed with 100 μL of ethanol solution (80% ethanol, 19.9% water and 0.1% formic acid). After 20 minutes at 4C, the samples were centrifuged at 15,000 g for 15 minutes and 180 μL of the supernatant was extracted and evaporated under

nitrogen gas. Samples were reconstituted in 60 µL HPLC-grade water and analyzed.

Cell extract analysis: cell contents were extracted using 1 mL of a methanol solution (80% methanol, 20% water) at dry ice temperature, after aspirating all media and quickly washing samples with ice-cold PBS. To prepare extract samples for LC-MS/MS analysis, the cell suspension was centrifuged at 14,000 g for 10 minutes and 100 µL of the supernatant was removed and evaporated under nitrogen gas. Samples were reconstituted in 60 µL HPLC-grade water and analyzed.

To evaluate the efficacy of respiratory chain inhibitors at the concentrations used, oxygen consumption rate was measured few minutes before and after addition of inhibitors or DMSO using a dedicated instrument (Seahorse Bioscience) as previously described (22). The change in oxygen consumption was calculated from the ratio of the two measurements.

Cell viability was measured with a calcein assay (Invitrogen C34852) after 24 hours incubation with respiratory chain inhibitors or DMSO.

Human Subjects

The criterion for inclusion of RCD patients in the study was a confirmed diagnosis of respiratory chain disease, based either on the presence of a pathogenic mutation or on reduced respiratory chain activity in muscle (Table 1). Patients were tested during an outpatient visit to the mitochondrial clinic at Massachusetts General Hospital. 17 blood samples were obtained from 16 patients (one patient tested on two separate visits 6 months apart). Control subjects were healthy adults participating in an epidemiologic study at Massachusetts Institute of Technology, selected at random balancing gender. The control and RCD groups were not matched for age, diet and medications. The study was approved by the Institutional Review Boards of Massachusetts

General Hospital and Massachusetts Institute of Technology, and informed consent was obtained from all subjects.

Plasma extraction and preparation for LC-MS/MS analysis was done as previously described (12). Briefly, venous blood was drawn into an EDTA-coated tube, centrifuged for 10 minutes at 2,000 g and 6C, and plasma was extracted. 165 µL plasma was mixed with 250 µL ethanol solution (80% ethanol, 19.9% water and 0.1% formic acid). After 2 hours at 4C, the samples were centrifuged at 15,000 g for 15 minutes, and 300 µL of the supernatant was extracted and evaporated under nitrogen. Samples were reconstituted in 60 µL HPLC-grade water and analyzed.

Metabolic Profiling

Metabolite levels were measured as previously described (12). Briefly, samples were separated on three different HPLC columns connected in parallel to a triple quadrupole mass spectrometer (4000 QTRAP, Applied Biosystems) operated in selected reaction monitoring mode. Each metabolite was identified by a combination of chromatographic retention time, precursor ion mass, and product ion mass. Metabolite quantification was performed by integrating the peak areas of product ions using MultiQuant software (Version 1.1, Applied Biosystems).

Statistical Analysis

The significance of univariate differences between cell culture conditions or between healthy and RCD subjects was evaluated using Wilcoxon rank sum test with $\alpha=0.05$. Multivariate linear regression was used to calculate the significance of a metabolite difference between patients and controls adjusting for covariates such as gender.

Classification

Classification of human subjects based on plasma metabolite levels was done with the support vector machine algorithm (SVM) using a linear kernel (Matlab Bioinformatics Toolbox version 3.1, The MathWorks). SVM is designed to identify a hyperplane that optimally separates two classes, in this case RCD and Healthy. SVM does not involve inversion of a covariance matrix, and is therefore unaffected by metabolite co-linearity and does not require dimensionality reduction. The linear kernel was chosen for simplicity.

Classification performance was estimated based on 100 rounds of cross-validation. In each round the training set included half the RCD samples and half the Healthy samples selected by random, and the test set included the remaining samples. Sensitivity, specificity and area under the receiver operating characteristic (ROC) curve were calculated based on the test set in each round of cross-validation, and averaged over all rounds. All classifiers were evaluated on the same collection of 100 random divisions to training and test parts.

In the comparison of model-derived metabolites to random metabolite sets (Figure 2B), the distribution of pairwise correlation coefficients amongst the subset of model-derived metabolites was typical to the superset of metabolites detected in plasma.

Acknowledgements

We thank N. Slate for assistance in coordination of the clinical study; A. Souza, R. Wei and E. Yang for assistance in mass spectrometry; and V. Gohil, M. Jain, T. Kitami and R. Nilsson for valuable discussions and comments on the manuscript. O.S was supported by a training grant for bioinformatics and integrative genomics from the National Human Genome Research Institute and a

fellowship from the Eli and Dorothy Berman Fund. V.K.M. was supported by an Early Career Physician-Scientist Award from the Howard Hughes Medical Institute. Clinical studies were conducted in part at the Clinical Research Center at the Massachusetts Institute of Technology funded by a grant RR-01066 from the National Center for Research Resources, National Institutes of Health. This work was supported by a grant from the Broad Institute Scientific Planning and Allocation of Resources Committee and a grant R01DK081457 from the National Institutes of Health to V.K.M.

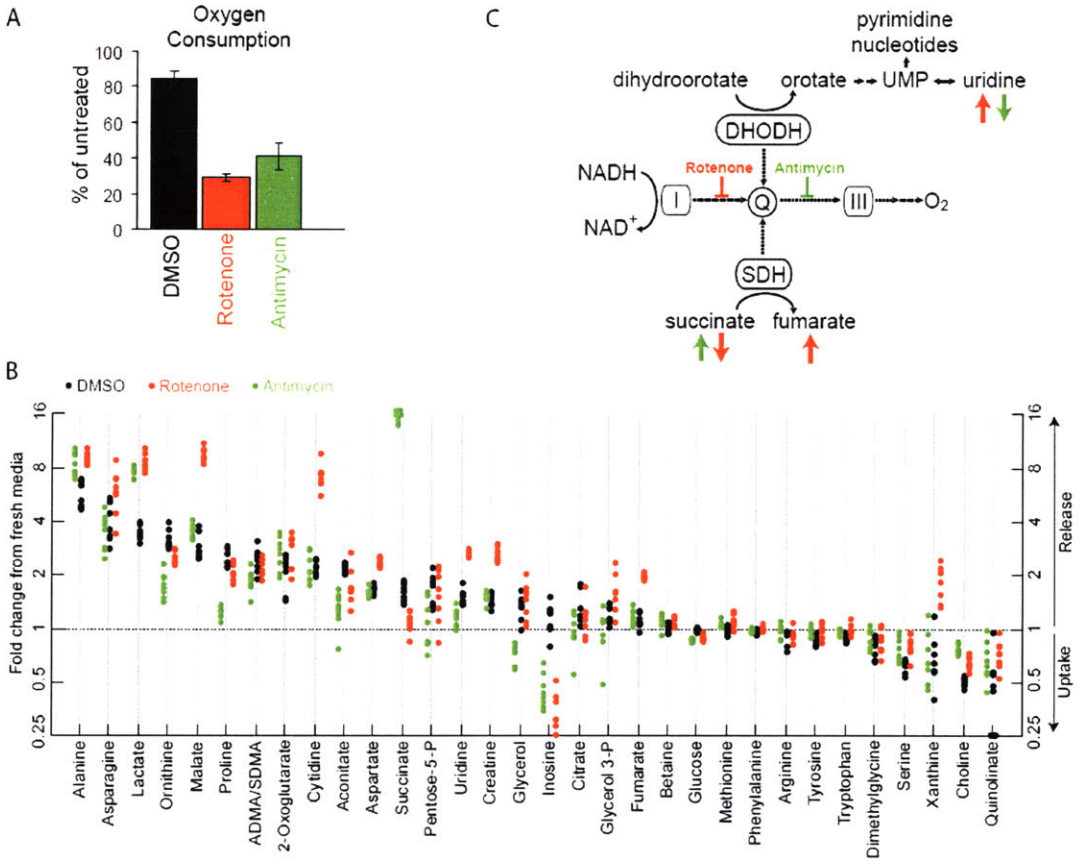


Figure 1: Metabolic profile of respiratory chain dysfunction in culture media.

A) Decrease in oxygen consumption following respiratory chain inhibition with rotenone and antimycin. Oxygen consumption was measured minutes before and after addition of inhibitors. Data presented as mean ± s.d. (n=3).

B) 32 metabolites significantly ($p<0.05$) altered with rotenone, antimycin or both inhibitors. Metabolites are ordered by the degree of change from fresh media in the control (DMSO) condition. Each condition was tested in 8 biological replicates.

C) Discordance between complex I and complex III profiles in pathways coupled to the respiratory chain. Dashed arrows represent electron flow. Red and green arrows denote trends with rotenone and antimycin, respectively. Abbreviations: I, complex I; III, complex III; SDH, succinate dehydrogenase (complex II); DHODH, dihydroorotate dehydrogenase; Q, conenzyme Q; UMP, uridine monophosphate.

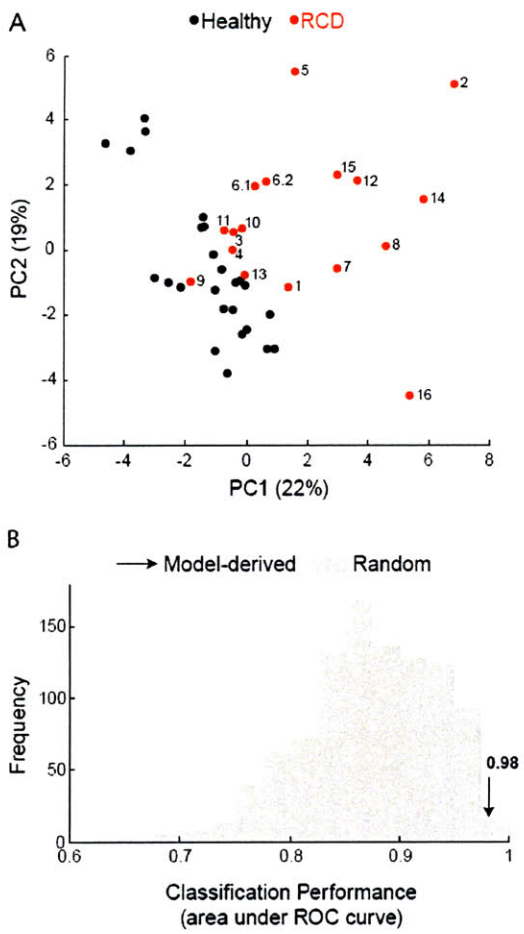


Figure 2: Separation between patients and controls using a metabolic profile derived from a cellular disease model.

A) Principal component analysis (unsupervised) of plasma samples from RCD patients (n=17) and healthy controls (n=25), based on metabolites altered in spent media of the cellular model (Figure 1B). Numbers near red markers denote patient identifiers (referencing Table 1). 6.1 and 6.2 represent two plasma samples from patient #6 obtained six months apart.

B) Comparison of the model-derived set of 29 metabolites (arrow) to randomly-selected sets of the same size. 1000 sets were selected randomly from all 87 metabolites detected in plasma. Multivariate linear classification using the support vector machine algorithm with cross-validation was used to estimate the classification performance, expressed as the area under the receiver operating characteristic (ROC) curve.

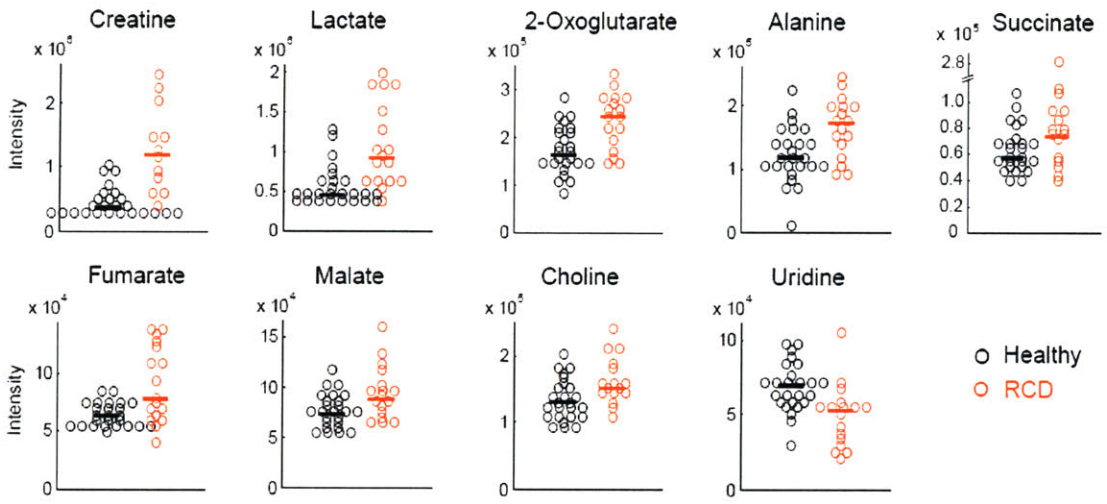


Figure 3: Plasma markers of human RCD.

Nine metabolites that significantly ($p<0.05$) differ between RCD and healthy subjects, and are altered in the same direction as in the cellular model (Figure 1B). Horizontal lines represent group medians.

	Gender	Age (yrs)	Clinical Presentation	Biochemical Features	Molecular Diagnosis
1	M	57	Chronic progressive external ophthalmoplegia and myopathy	c. I,III,IV, RRF, SDH+/COX- (muscle)	Large mtDNA deletion
2	M	54	Exercise intolerance, myopathy, myoclonus, neuropathy, diabetes, dysphagia, lipomatosis and SNHL	Lactic acidemia	mt A8344G
3	F	46	Myopathy, neuropathic pain, exercise intolerance and stroke	c. I (muscle), elevated CK	
4	M	62	Encephalomyopathy	c. I,II,III,IV (muscle)	Multiple mtDNA deletions
5	F	25	Muscle cramps, exercise intolerance, TIA and complex migraine	Lactic acidemia	mt A3243G
6	F	52	Cerebellar ataxia, dysarthria, peripheral neuropathy and SNHL	c. I,II,IV (muscle)	
7	F	17	Headaches, GI dysmotility, hypotonia, muscle cramps	c. I (muscle)	
8	F	44	TIA, complex migraine, Crohn's disease, hearing loss, muscle weakness and cramps, exercise intolerance.	c. III (muscle)	
9	F	46	Exercise intolerance, dystonia	c. IV (muscle)	
10	M	29	Myopathy, chronic headache, sleep apnea, exercise intolerance, GI dysmotility	c. I,III (muscle)	
11	F	35	Neurodegenerative encephalopathy with spasticity and involuntary movements	c. I (muscle), lactic acidemia, elevated CK	
12	F	51	Stroke, migraine, SNHL, expressive aphasia	Elevated CK	mt A3243G
13	M	42	Myopathy, chronic fatigue, exercise intolerance and hemochromatosis	c. III,IV (muscle)	
14	F	54	Exercise intolerance, GI dysmotility		mt A3243G
15	M	22	Seizures, myoclonus, cardiomyopathy, mild dementia	Lactic acidemia	mt A3243G
16	M	51	Myopathy, exercise intolerance, myalgias with fasciculations, SNHL, complex migraine and atrial fibrillation	c. I,III,IV (muscle), elevated CK	mtDNA depletion

Table 1: Demographic, clinical, biochemical and molecular characteristics of RCD Patients.

Abbreviations: SNHL, sensorineural hearing loss; TIA, transient ischemic attack; RRF, ragged-red fibers; SDH+/COX-, succinate dehydrogenase positive / cytochrome C oxidase negative; CK, creatine kinase.

	Lactate, Alanine	Lactate, Alanine + TCA Cycle Intermediates	All 9 Metabolites
Area Under ROC Curve	0.82	0.92	0.96
Sensitivity[1]	47%	73%	87%
Specificity[2]	38%	57%	73%

[1] Sensitivity at 100% specificity
[2] Specificity at 100% sensitivity

Table 2: Multivariate classification of plasma from RCD patients and healthy controls.

Comparison of classifiers based on 3 subsets of the 9 plasma markers identified (Figure 3). ROC, receiver operating characteristic.

References

1. Uusimaa, J., *et al.* (2007) Prevalence, segregation, and phenotype of the mitochondrial DNA 3243A>G mutation in children. *Ann Neurol* 62:278-287.
2. Thorburn, D. R. (2004) Mitochondrial disorders: prevalence, myths and advances. *J Inherit Metab Dis* 27:349-362.
3. Bernier, F. P., *et al.* (2002) Diagnostic criteria for respiratory chain disorders in adults and children. *Neurology* 59:1406-1411.
4. Morava, E., *et al.* (2006) Mitochondrial disease criteria: diagnostic applications in children. *Neurology* 67:1823-1826.
5. Walker, U. A., Collins, S., & Byrne, E. (1996) Respiratory chain encephalomyopathies: a diagnostic classification. *Eur Neurol* 36:260-267.
6. Barshop, B. A. (2004) Metabolomic approaches to mitochondrial disease: correlation of urine organic acids. *Mitochondrion* 4:521-527.
7. Shoffner, J. M. (1999) Oxidative phosphorylation disease diagnosis. *Semin Neurol* 19:341-351.
8. Haas, R. H., *et al.* (2008) The in-depth evaluation of suspected mitochondrial disease. *Mol Genet Metab* 94:16-37.
9. Haas, R. H., *et al.* (2007) Mitochondrial disease: a practical approach for primary care physicians. *Pediatrics* 120:1326-1333.
10. Allen, J., *et al.* (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* 21:692-696.
11. Lewis, G. D., *et al.* (2008) Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J Clin Invest* 118:3503-3512.
12. Shaham, O., *et al.* (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Mol Syst Biol* 4:214.

13. Gutman, M. & Silman, N. (1972) Mutual inhibition between NADH oxidase and succinoxidase activities in respiring submitochondrial particles. *FEBS Lett* 26:207-210.
14. Jones, M. E. (1980) Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu Rev Biochem* 49:253-279.
15. Loffler, M., Jockel, J., Schuster, G., & Becker, C. (1997) Dihydroorotat-ubiquinone oxidoreductase links mitochondria in the biosynthesis of pyrimidine nucleotides. *Mol Cell Biochem* 174:125-129.
16. Arnold, D. L., Taylor, D. J., & Radda, G. K. (1985) Investigation of human mitochondrial myopathies by phosphorus magnetic resonance spectroscopy. *Ann Neurol* 18:189-196.
17. King, M. P. & Attardi, G. (1989) Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* 246:500-503.
18. Webster, D. R., Becroft, D. M. O., van Gennip, A. H., & Van Kuilenburg, A. B. P. (2001) in *Metabolic and Molecular Bases of Inherited Disease*, ed. Scriver, C. (McGraw-Hill, New York).
19. Atkuri, K. R., *et al.* (2009) Inherited disorders affecting mitochondrial function are associated with glutathione deficiency and hypocitrullinemia. *Proc Natl Acad Sci U S A* 106:3941-3945.
20. DiMauro, S. & Mancuso, M. (2007) Mitochondrial diseases: therapeutic approaches. *Biosci Rep* 27:125-137.
21. Falk, M. J., *et al.* (2008) Metabolic pathway profiling of mitochondrial respiratory chain mutants in C. elegans. *Mol Genet Metab* 93:388-397.
22. Wu, M., *et al.* (2007) Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. *Am J Physiol Cell Physiol* 292:C125-136.

Chapter 5

—

Implications and Future Directions

Implications and Future Directions

In 2005 when the metabolic profiling initiative was launched at Broad Institute, there were few research papers reporting the application of this technology to mammalian systems. Today, 4 years later, metabolic profiling is a much better appreciated tool for biomedical research, contributing to the study of various diseases including cancer (1), cardiovascular disease (2), infectious diseases (3), diabetes (4-7) and mitochondrial disease. This dissertation demonstrates the application of metabolic profiling, particularly mass spectrometry-based, to the study of human physiology and disease. We discuss here the main implications of this work and some future directions.

Implications

A Multidimensional View of Insulin Sensitivity

Insulin sensitivity is typically defined by the effect of insulin on blood glucose levels. Our work on the human response to a glucose challenge has revealed that other axes of insulin action, particularly the suppression of proteolysis and lipolysis, contain orthogonal information. We postulate that integrating this information in a multidimensional insulin response profile could help better characterize insulin resistance in pre-diabetic individuals.

We detected a particularly strong correlation between insulin resistance and the blunted decrease of branched-chain amino acids (BCAA) following a glucose challenge. BCAA elevations in obesity, a condition associated with insulin resistance, have been reported previously (8), and a recent report of metabolic profiling in obese and lean humans has confirmed this connection (5). Physiological effects of BCAA have been noted in various aspects of energy metabolism, including the stimulation of insulin secretion in pancreatic beta cells

(9) and the control of appetite via the mTOR pathway in hypothalamic neurons (10). It is therefore likely that in the future BCAA will be used clinically as a marker of insulin resistance.

Novel Findings in the Human Response to a Glucose Challenge

Taking a global profiling approach, we identified physiological effects of glucose ingestion that were previously unappreciated. These included a striking increase in plasma bile acids, as well as decreases in the urea cycle intermediates ornithine and citrulline and in the purine base hypoxanthine. The strong and sustained elevation of bile acids in response to glucose alone was surprising, given that fat and protein ingestion are known as the major stimuli for gallbladder contraction. Since the publication of these results, the elevation of plasma bile acids in response to glucose ingestion was validated by an independent group (7). Of particular interest is the reported role of bile acids in regulating energy metabolism, where they act as a signal for fat and muscle cells to increase their energy expenditure through activation of thyroid hormone (11). Taken together, bile acids could be playing a role in glucose homeostasis.

Media Profiling of a Cellular Disease Model for the Identification of a Disease Signature

In Chapter 4 we have demonstrated the utility of cell culture media profiling for defining a plasma signature of human mitochondrial disease. Media profiling of appropriate cellular disease models can help focus the subsequent clinical investigation on a smaller number of metabolites. This type of study design is particularly important in the clinical investigation of rare disorders, where the statistical power is limited by the relatively small number of patients, and where potential confounding factors may exist. This use of metabolic profiling technology could simplify the identification of disease markers and

facilitate understanding of pathologic processes in various other conditions, including common diseases.

A Plasma Signature of Mitochondrial Disease

This dissertation describes the first study of mitochondrial disease to evaluate such a large number of metabolites. In an unbiased fashion, we have confirmed previously suggested markers and have identified novel ones, providing insight into the pathogenesis of these disorders. We have detected plasma markers reflecting substrate oxidation (lactate, alanine and intermediates of the tricarboxylic acid cycle), pyrimidine biosynthesis (uridine), energy charge (creatine) and choline metabolism. We show that a combination of all these metabolites provides better classification of patients and controls compared to the currently used markers. Of particular note is the identification of low uridine levels, suggesting that uridine supplementation might be beneficial to some patients.

Future Directions

Clinical Interpretation of the Insulin Response Profile

In this dissertation we have proposed the concept of a multidimensional insulin response profile in pre-diabetic individuals. There is currently no information on the relationship between an individual's insulin response profile and disease progression. Prospective clinical studies, in which subjects are followed over several years, could help explain this relationship. Understanding which profiles are associated with higher risk of developing diabetes could improve the medical care of pre-diabetic patients, and delay or prevent the onset of diabetes.

Validation of the Plasma Signature of Mitochondrial Disease

We have identified here a plasma metabolic signature of mitochondrial disease. To test the reproducibility of this signature, and in particular the novel metabolic markers, we are working on the replication of our study in an independent cohort of patients. We also plan to evaluate the correlation of markers with disease severity. If such a correlation exists, the signature we have identified could be very valuable as a marker of disease progression or therapeutic response.

Physiologic perturbations are useful for unmasking metabolic abnormalities. The oral glucose tolerance test, for example, challenges the body's mechanisms of glucose homeostasis and can reveal glucose intolerance. In a similar fashion, provocative tests in mitochondrial disease patients can accentuate abnormalities that are difficult to detect at baseline. Exercise challenges the capacity of muscle mitochondria to oxidize fuel and produce ATP. In normal individuals, exceeding the mitochondrial capacity leads to elevated plasma lactate. Exercise also elevates muscle levels of succinate, fumarate and malate (12). We expect that exercise testing of mitochondrial disease patients will emphasize some of the abnormalities we have detected at baseline, and perhaps uncover additional ones.

The Relationship between Genotype and Metabolic Phenotype in Mitochondrial Dysfunction

Mitochondrial respiratory chain disorders are caused by dozens of different genes and hundreds of different mutations in either the mitochondrial or the nuclear DNA. The striking differences between complex I inhibition and complex III inhibition in the current study suggest that different genotypes could differ in their metabolic phenotype. An understanding of these differences would be tremendously helpful for the molecular diagnosis and care of patients. Such

genotype-specific profiles also hold potential for understanding the pathogenesis of mitochondrial disorders. Future well-powered clinical studies or studies of genetic models of mitochondrial disease in cells or animals could provide such insight.

References

1. Sreekumar, A., *et al.* (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457:910-914.
2. Lewis, G. D., *et al.* (2008) Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J Clin Invest* 118:3503-3512.
3. Munger, J., *et al.* (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol* 26:1179-1186.
4. Makinen, V. P., *et al.* (2008) 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* 4:167.
5. Newgard, C. B., *et al.* (2009) A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 9:311-326.
6. Shaham, O., *et al.* (2008) Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity. *Mol Syst Biol* 4:214.
7. Zhao, X., *et al.* (2009) Changes of the plasma metabolome during an oral glucose tolerance test: is there more than glucose to look at? *Am J Physiol Endocrinol Metab* 296:E384-393.
8. She, P., *et al.* (2007) Obesity-related elevations in plasma leucine are associated with alterations in enzymes involved in branched-chain amino acid metabolism. *Am J Physiol Endocrinol Metab* 293:E1552-1563.
9. Kalogeropoulou, D., Lafave, L., Schweim, K., Gannon, M. C., & Nuttall, F. Q. (2008) Leucine, when ingested with glucose, synergistically stimulates insulin secretion and lowers blood glucose. *Metabolism* 57:1747-1752.
10. Kahn, B. B. & Myers, M. G., Jr. (2006) mTOR tells the brain that the body is hungry. *Nat Med* 12:615-617.

11. Watanabe, M., *et al.* (2006) Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* 439:484-489.
12. Gibala, M. J., Tarnopolsky, M. A., & Graham, T. E. (1997) Tricarboxylic acid cycle intermediates in human muscle at rest and during prolonged cycling. *Am J Physiol* 272:E239-244.

Appendix A

—

Supplementary Material for Chapter 3

Supplementary Material:

Metabolic profiling of the human response to a glucose challenge reveals distinct axes of insulin sensitivity

Contents:

Item 1: HPLC method

Item 2: Mass spectrometry method

Item 3: Metabolite interferences

Item 4: Reproducibility of metabolite measurement in plasma

Item 5: Chemical classes of monitored metabolites

Item 6: Metabolites reported in the main text

Item 7: Metabolites detected in plasma

Item 8: Metabolites monitored but not detected in plasma

Item 1: HPLC method

Three different HPLC systems were used sequentially in time, and connected to the mass spectrometer in parallel. All columns were purchased from Phenomenex (Torrance, CA). The following table lists the parameters of each HPLC system:

	System 1	System 2	System 3
Mobile Phase	**A**[a]**:** 99.9% water 0.1% acetic acid **B**[b]**:** 99.9% acetonitrile 0.1% acetic acid	**A:** 79.75% water 20% acetonitrile 0.25% ammonium hydroxide 10 mM ammonium acetate **B:** 79.75% acetonitrile 20% water 0.25% ammonium hydroxide 10 mM ammonium acetate	**A:** 95% water 5% acetonitrile 5 mM ammonium acetate **B:** 95% acetonitrile 5% water 5 mM ammonium acetate
Column	Luna phenyl-hexyl (4.6x50 mm, 5µm)	Luna Amino (4.6x50 mm, 5µm)	Synergi Polar-RP (4.6x50 mm, 4µm)
Gradient	From 0%B and 1 mL/min to 90%B and 2 mL/min in 0.7 minutes	From 100%B and 1.5 mL/min to 0%B and 2.5 mL/min in 1.6 minutes	From 5%B and 1 mL/min to 95%B and 2 mL/min in 2.65 minutes
Total Run Time (minutes)	1.4	3.4	3.75
Injection Volume (µL)	5	10	10

[a] A: Aqueous phase
[b] B: Organic phase.

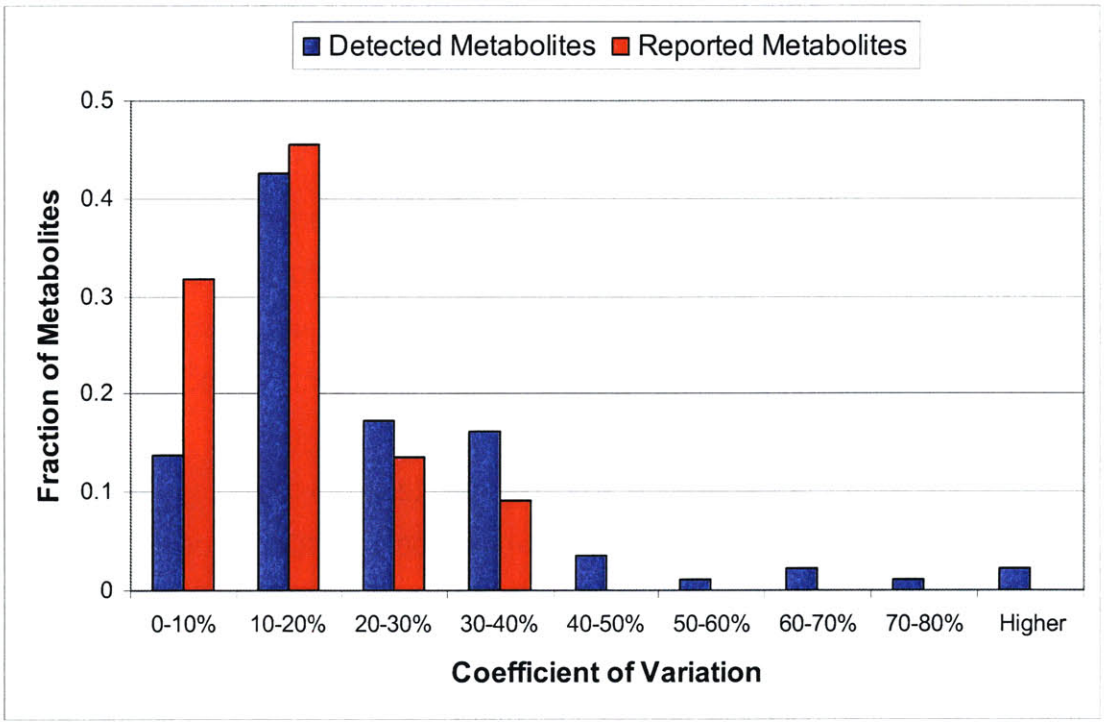
Item 2: Mass spectrometry method

A Turbo electrospray ionization source was used. The ion spray potentials were 5,000 volt in the positive mode and 4,200 volt in the negative mode. Zero air was used for the nebulizer and bath gases, and N_2 was used for the curtain and collision gases. The gas pressures used were 50 psi for the nebulizer gas, 60 psi for the bath gas, 20 psi for the curtain gas and 7 psi for the collision gas. The bath gas temperature was 400 °C.

Item 3: Metabolite interferences

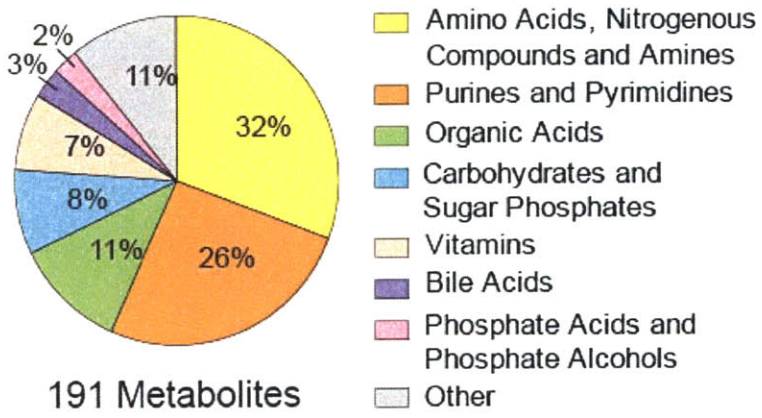
The HPLC-MS/MS method was unable to distinguish between a few of the reported metabolites and other tested metabolites, due to a combination of isobaric overlap and insufficient chromatographic resolution. Leucine and isoleucine were indistinguishable, and are therefore always mentioned together in the text. In the other instances of sets of indistinguishable metabolites, one of the metabolites in the set was likely to be present in the samples in much higher concentrations than the rest of the set, based on published concentrations in human plasma and given the nature of an oral glucose tolerance test. In these instances, it was assumed that the effect of the non-prevalent metabolites on the measurement was negligible, and only the prevalent metabolite was reported in the main text. The sets of indistinguishable metabolites involving a reported metabolite are listed below, with the prevalent metabolite first: {glucose, galactose, fructose}, {β-hydroxybutyrate, malonate}, and {valine, guanidinoacetate}.

Item 4: Reproducibility of metabolite measurement in plasma



The variation associated with plasma sample preparation and LC-MS/MS analysis was assessed by splitting one plasma sample into 7 parts and processing each part separately. The coefficient of variation (CV) equals the standard deviation of metabolite level divided by the mean metabolite level. The CV distribution of metabolites reported in the main text (red bars, Item 6 below) is compared to the distribution of all detected metabolites (blue bars, Item 7 below).

Item 5: Chemical classes of monitored metabolites



The chemical classification is based on the chemical taxonomy annotation in the Human Metabolome Database (Wishart, D.S. et al. HMDB: the Human Metabolome Database. Nucleic Acids Res 35, D521-526 (2007).

Item 6: Metabolites reported in the main text

Metabolite Name	Q1[a]	Q3[b]	DP[c]	CE[d]	IP[e]	HPLC[f]	Standard Source	Standard Catalog Number
Alanine	90.0	44.0	25	15	+	1	Sigma[g]	A-7627
Arginine	175.1	70.0	25	30	+	1	Sigma	A-5131
-hydroxybutyrate	103.0	59.0	-40	-15	-	3	Sigma	54920
Citrulline	174.1	131.0	-50	-15	-	3	Sigma	27510
Glucose	179.1	89.0	-50	-15	-	2	Sigma	49159
Glycerol	93.0	57.0	20	21	+	1	Shelton Scientific - IBI	IB15762
Glycochenodeoxycholic acid	448.3	74.0	-80	-60	-	3	Sigma	G0759
Glycocholic acid	464.3	74.0	-30	-60	-	3	Sigma	G2878
Hippuric acid	178.1	134.0	-50	-16	-	3	Sigma	112003
Histidine	156.1	110.0	25	23	+	1	Sigma	H-8125
Hypoxanthine	135.0	92.0	-50	-23	-	3	Sigma	56700
Isoleucine	132.1	86.0	50	20	+	1	Sigma	I-2752
Lactate	89.0	43.0	-40	-20	-	2	Sigma	69771
Leucine	132.1	86.0	50	20	+	1	Sigma	L-8000
Lysine	147.1	84.0	25	25	+	1	Sigma	G-3126
Malate	133.0	115.0	-40	-20	-	3	Sigma	M-0750
Methionine	150.1	61.0	40	30	+	1	Sigma	M-9625
Ornithine	133.1	70.0	40	30	+	1	Sigma	75480
Phenylalanine	166.1	120.0	50	17	+	1	Sigma	P-2126
Pyruvate	87.0	43.0	-30	-12	-	3	Sigma	107360
Taurochenodeoxycholic acid	498.3	80.0	-90	-90	-	3	Sigma	86335
Tyrosine	182.1	136.5	25	17	+	1	Sigma	T-3754
Valine	118.1	72.0	25	20	+	1	Sigma	V-0500
Xanthine	151.0	108.0	-40	-30	-	2	Sigma	95490

[a] Q1: precursor ion mass, in daltons (Da)
[b] Q3: product ion mass, in daltons (Da)
[c] DP: De-clustering Potential, in volts (V)
[d] CE: Collision Energy, in volts (V)
[e] IP: Ionization Polarity
[f] HPLC: the HPLC system in which the metabolite was measured. See the HPLC section (Item 1 above) for the parameters of each system.
[g] Sigma: Sigma-Aldrich Co.

Item 7: Metabolites detected in plasma (metabolite #1-97)

Column description:

KEGG: Compound ID from the Kyoto Encyclopedia of Genes and Genomes, http://www.genome.ad.jp

Q1: precursor ion mass, in daltons (Da)

Q3: product ion mass, in daltons (Da)

HPLC: the HPLC system in which the metabolite was measured

CV: Coefficient of variation of plasma sample analysis

The name used for a metabolite in the main text is in square brackets (if different from KEGG)

#	KEGG	Name	Q1	Q3	HPLC	CV
1	C01089	(R)-3-Hydroxybutanoate [β-hydroxybutyrate]	103.0	59.0	3	27
2	C00186	(S)-Lactate [Lactate]	89.0	43.0	2	7
3	C00149	(S)-Malate [Malate]	133.0	115.0	3	32
4	C00026/ C06104	2-Oxoglutarate/Adipate	145.1	101.0	3	20
5	C00632	3-Hydroxyanthranilate	154.0	136.2	1	54
6	C00197	3-Phospho-D-glycerate	185.0	97.0	2	78
7	C01015	4-Hydroxy-L-proline	132.1	86.2	1	15
8	C00847	4-Pyridoxate	182.1	138.0	3	15
9	C01996	Acetylcholine	146.1	87.0	1	18
10	C00008	ADP	426.0	79.0	2	66
11	C01551	Allantoin	159.0	116.0	1	13
12	C00020	AMP	346.1	79.0	2	85
13	C00108	Anthranilate	138.0	120.0	1	18
14	C00002	ATP	506.0	159.0	3	34
15	C00719	Betaine	118.1	58.0	1	14
16	C00486	Bilirubin	583.3	285.0	3	69
17	C00114	Choline	104.1	60.0	1	18
18	C00417	cis-Aconitate	173.0	129.0	3	32
19	C00158/ C00311	Citrate/Isocitrate	191.0	111.0	3	
20	C00300	Creatine	132.1	90.0	1	16
21	C00791	Creatinine	114.1	44.0	1	14
22	C01094/ C00085/ C00103/ C00668	D-Fructose 6-phosphate/D-Glucose 1-phosphate/alpha-D-Glucose 6-phosphate/D-Fructose 1-phosphate	259.0	97.0	2	18
23	C00191	D-Glucuronate	193.0	113.0	3	27
24	C00117/ C00199	D-Ribose 5-phosphate/D-Ribulose 5-phosphate	229.0	97.0	3	33
25	C00365	dUMP	307.0	79.0	3	29
26	C00122/ C01384	Fumarate/Maleate	115.0	71.0	2	18
27	C00035	GDP	442.0	79.0	3	34

28	C00341	Geranyl diphosphate	313.1	79.1	3	46
29	C00031/ C00124	Glucose	179.1	89.0	2	8
30	C00116	Glycerol	93.0	57.0	1	5
31	C00111	Glycerone phosphate	169.0	97.0	3	35
32	C05466	Glycochenodeoxycholate [GCDCA]	448.3	74.0	3	16
33	C01921	Glycocholate [GCA]	464.3	74.0	3	15
34	C00044	GTP	522.0	159.0	3	
35	C01586	Hippurate [Hippuric acid]	178.1	134.0	3	24
36	C00544	Homogentisate	167.0	123.0	3	30
37	C00262	Hypoxanthine	135.0	92.0	3	19
38	C00130	IMP	347.0	79.0	2	18
39	C00294	Inosine	267.1	135.0	3	30
40	C00956	L-2-Aminoadipate	160.1	116.0	2	33
41	C00243	Lactose	341.1	161.0	2	8
42	C00041	L-Alanine [Alanine]	90.0	44.0	1	16
43	C00062	L-Arginine [Arginine]	175.1	70.0	1	13
44	C00152	L-Asparagine	133.1	74.0	1	14
45	C00049	L-Aspartate	134.0	74.0	1	15
46	C00327	L-Citrulline [Citrulline]	174.1	131.0	3	30
47	C00097	L-Cysteine	122.0	76.0	1	27
48	C00407/ C00123	Leucine/Isoleucine	132.1	86.2	1	8
49	C00025	L-Glutamate	148.1	84.0	1	19
50	C00064	L-Glutamine	147.1	84.0	1	14
51	C00135	L-Histidine [Histidine]	156.1	110.0	1	14
52	C00155	L-Homocysteine	136.0	90.0	1	17
53	C00328	L-Kynurenine	207.1	144.0	3	16
54	C00047	L-Lysine [Lysine]	147.1	84.0	1	15
55	C00073	L-Methionine [Methionine]	150.1	61.0	1	34
56	C00077	L-Ornithine [Ornithine]	133.1	70.0	1	17
57	C00079	L-Phenylalanine [Phenylalanine]	166.1	120.2	1	6
58	C00148	L-Proline	116.1	70.0	1	11
59	C00065	L-Serine	106.0	60.0	1	19
60	C00188	L-Threonine	120.1	74.0	1	18
61	C00078	L-Tryptophan	205.1	188.3	1	3
62	C00082	L-Tyrosine [Tyrosine]	182.1	136.3	1	7
63	C02170	Methylmalonate	117.0	73.0	2	
64	C03406	N-(L-Arginino)succinate	291.1	70.0	1	46
65	C01026	N,N-Dimethylglycine	104.1	58.0	1	19
66	C00003	NAD+	662.1	540.0	2	
67	C03626	NG,NG-Dimethyl-L-arginine (asymmetric)	201.1	156.0	3	36
68	C03626	NG,NG-Dimethyl-L-arginine (asymmetric/symmetric)	203.1	70.3	1	36
69	Not found	NG,NG-Dimethyl-L-arginine (symmetric)	201.1	131.0	3	
70	C03884	Ngamma-Monomethyl-L-arginine	189.1	70.0	1	
71	C00153	Nicotinamide	123.0	80.0	1	13
72	C00295	Orotate	155.0	111.0	3	31
73	C01103	Orotidine 5'-phosphate	367.0	323.1	2	
74	C00209	Oxalate	89.0	61.0	3	33
75	C00036	Oxaloacetate	131.0	87.0	3	24

76	C00127	Oxidized glutathione	611.2	306.0	2	
77	C00864	Pantothenate	218.1	88.0	3	46
78	C00166	Phenylpyruvate	163.0	91.0	2	29
79	C03722	Pyridine-2,3-dicarboxylate	166.0	122.0	2	25
80	C00314	Pyridoxine	170.1	152.2	1	7
81	C00022	Pyruvate	87.0	43.0	3	
82	C00780	Serotonin	177.1	160.0	1	25
83	C00093	sn-Glycerol 3-phosphate	171.0	79.0	2	9
84	C00042	Succinate	117.0	73.0	3	33
85	C00245	Taurine	126.0	108.0	1	14
86	C05465	Taurochenodeoxycholate [TCDCA]	498.3	80.0	3	17
87	C00068	Thiamin diphosphate	423.0	303.2	2	9
88	C01104	Trimethylamine N-oxide	76.1	58.0	1	19
89	C00015	UDP	403.0	79.0	3	36
90	C00029/ C00052	UDPglucose/UDP-D-galactose	565.1	323.0	2	132
91	Not found	Unidentified peak	149.0	87.0	3	26
92	C00106	Uracil	111.0	42.0	2	24
93	C00366	Urate	167.0	124.0	3	24
94	C00299	Uridine	243.1	110.0	2	16
95	C00075	UTP	483.0	159.0	3	
96	C00183/ C00581	Valine	118.1	72.0	1	7
97	C00385	Xanthine	151.0	108.0	2	11

Item 8: Metabolites monitored, not detected in plasma (metabolite #98-191)

Column description:

KEGG: Compound ID from the Kyoto Encyclopedia of Genes and Genomes, http://www.genome.ad.jp

Q1: precursor ion mass, in daltons (Da)

Q3: product ion mass, in daltons (Da)

HPLC: the HPLC system in which the metabolite was measured

#	KEGG	Name	Q1	Q3	HPLC
98	C00631	2-Phospho-D-glycerate	185.0	79.1	3
99	C01179	3-(4-Hydroxyphenyl)pyruvate	179.0	107.0	3
100	C00575	3',5'-Cyclic AMP	328.1	134.0	2
101	C00942	3',5'-Cyclic GMP	344.0	150.0	3
102	C05145	3-Aminoisobutanoate	104.1	86.0	1
103	C03227	3-Hydroxy-L-kynurenine	223.1	162.0	3
104	C05594	3-Methoxy-4-hydroxyphenylethyleneglyco	183.1	150.0	2
105	Not found	3-Nitro-L-tyrosine	227.1	181.3	1
106	C02642	3-Ureidopropionate	133.1	115.0	1
107	C00334	4-Aminobutanoate	104.1	87.0	1
108	C00156	4-Hydroxybenzoate	137.0	93.0	3
109	C03479	5-Formyltetrahydrofolate	472.2	315.0	2
110	C05635	5-Hydroxyindoleacetate	190.1	146.0	2
111	C00643	5-Hydroxy-L-tryptophan	221.1	204.0	1
112	C00440	5-Methyltetrahydrofolate	458.2	329.0	2
113	C00164	Acetoacetate	101.0	57.0	2
114	C00024	Acetyl-CoA	403.6	79.0	3
115	C00147	Adenine	134.1	107.0	2
116	C00212	Adenosine	266.1	134.0	2
117	C03794	Adenylosuccinate	462.1	79.0	3
118	C00072	Ascorbate	175.0	115.0	3
119	C00120	Biotin	243.1	200.0	2
120	C00487	Carnitine	163.1	85.0	1
121	C00386	Carnosine	227.1	110.0	1
122	C00112	CDP	402.0	158.9	3
123	C02528	Chenodeoxycholate	391.3	373.5	2
124	C00695	Cholate	407.3	343.0	3
125	C00055	CMP	322.1	79.0	2
126	Not found	Cotinine	177.1	80.0	1
127	C00063	CTP	482.0	159.0	3
128	C02823	Cyanocobalamin	678.3	147.3	1
129	Not found	Cystamine	153.0	108.0	1
130	C01678	Cysteamine	78.0	61.0	1
131	C00475	Cytidine	244.1	112.0	1

132	C00380/ C00388	Cytosine/Histamine	112.0	95.0	1
133	C00705	dCDP	386.0	79.0	3
134	C00239	dCMP	306.1	79.0	3
135	C00458	dCTP	466.0	159.0	3
136	C00559	Deoxyadenosine	252.1	136.3	1
137	C00881	Deoxycytidine	228.1	73.0	1
138	C00526	Deoxyuridine	229.1	113.0	1
139	C00095	D-Fructose	179.1	89.0	2
140	C05378/ C00665/ C00660	D-Fructose 1,6-bisphosphate/D-Fructose 2,6-bisphosphate/D-Glucose 1,6-bisphosphate	339.0	79.0	3
141	C00577	D-Glyceraldehyde	89.0	59.0	2
142	C00415	Dihydrofolate	442.2	176.0	2
143	C03758	Dopamine	154.1	137.2	1
144	C00794	D-Sorbitol	181.1	89.0	2
145	C00364	dTMP	321.1	195.0	3
146	C00459	dTTP	481.0	383.0	3
147	C00460	dUTP	467.0	370.0	3
148	C05589	Epinephrine/Normetanephrine	184.1	166.3	1
149	C00346	Ethanolamine phosphate	142.0	44.0	1
150	C00504	Folate	440.1	311.0	3
151	C00051	Glutathione	306.1	143.0	2
152	C00037	Glycine	76.0	30.5	1
153	C00144	GMP	362.1	79.0	3
154	C00242	Guanine	150.0	133.0	2
155	C00387	Guanosine	282.1	150.1	2
156	C01817	Homocystine	267.1	132.0	3
157	C05582	Homovanillate	181.1	137.1	2
158	C01717	Kynurenate	188.0	144.0	2
159	C01724	Lanosterol	425.4	212.1	2
160	C02291	L-Cystathionine	221.1	134.0	3
161	C00263	L-Homoserine	120.1	74.2	1
162	C05588	L-Metanephrine	198.1	180.5	1
163	C05589	L-Normetanephrine	184.1	134.0	1
164	C00083	Malonyl-CoA	425.6	79.0	3
165	C01598	Melatonin	233.1	174.3	1
166	Not found	Methylhydroxyisobutyrate	119.1	87.0	1
167	C00137	myo-Inositol	179.1	81.0	2
168	C00004	NADH	664.1	408.0	2
169	C00006	NADP+	742.1	620.0	2
170	C00005	NADPH	744.1	408.0	2
171	C05926	Neopterin	252.1	192.0	2
172	C00253	Nicotinate	122.0	78.0	3
173	C01185	Nicotinate D-ribonucleotide	334.0	290.0	2
174	C05127	N-Methylhistamine	126.1	109.0	1
175	C00074	Phosphoenolpyruvate	167.0	79.0	2
176	C00163	Propanoate	73.0	55.0	3
177	C00584	Prostaglandin E2	351.2	315.3	3
178	C00018	Pyridoxal phosphate	246.0	97.0	3
179	C00021	S-Adenosyl-L-homocysteine	385.1	136.3	1

180	C07588	Salicyluric acid	194.1	150.0	3
181	C00315	Spermidine	146.2	72.0	1
182	C00750	Spermine	203.2	129.3	1
183	C00089	Sucrose	341.1	179.0	3
184	C05122	Taurocholate	514.3	123.8	3
185	C00214	Thymidine	243.1	127.0	1
186	C00178	Thymine	125.0	42.0	2
187	C00167	UDPglucuronate	579.0	403.0	2
188	C00105	UMP	323.0	79.0	3
189	C01762	Xanthosine	285.1	153.0	1
190	C00655	Xanthosine 5'-phosphate	363.0	79.0	3
191	C02470	Xanthurenic acid	204.0	160.0	2

Appendix B

—

Supplementary Material for Chapter 4

Supplementary Material:

A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent cell culture media

Contents:

Figure S1

Table S1

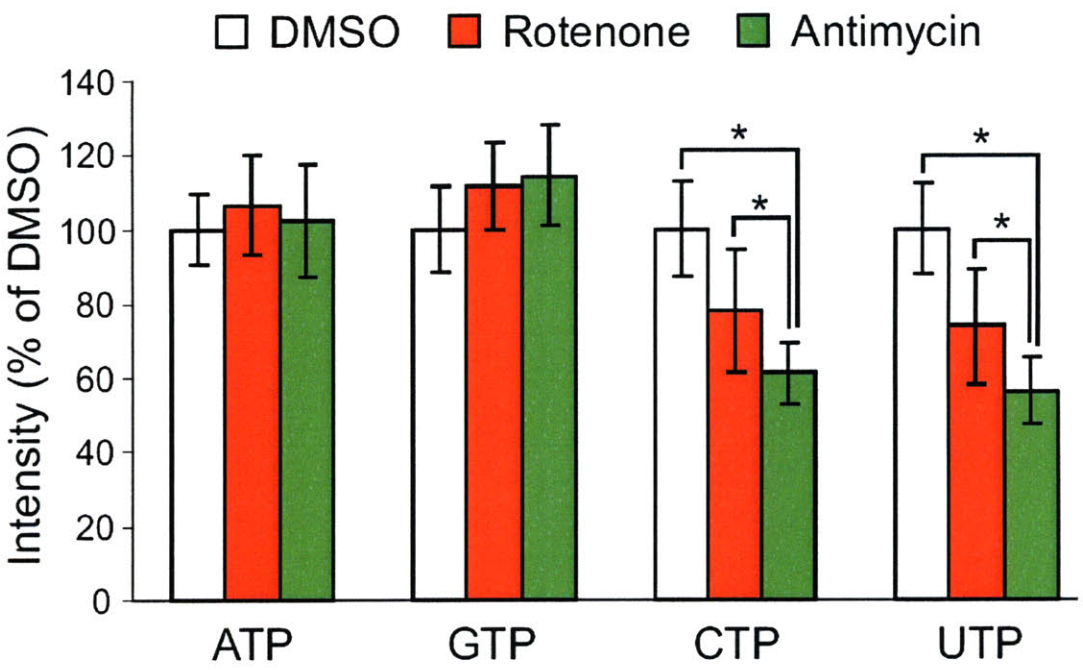


Figure S1: Intracellular levels of nucleotides in myotubes treated with RC inhibitors.

Levels of purine and pyrimidine nucleotides were measured in cell extract following 8 hours of inhibition with rotenone or antimycin, and compared to treatment with vehicle (DMSO). Data is presented as mean ± s.d. over 8 biological replicates per condition.

* denotes p-value < 0.05.

Abbreviations: ATP, adenine triphosphate; GTP, guanine triphosphate; CTP, cytidine triphosphate; UTP, uridine triphosphate.

Metabolite	% of Total Weight	Sign
Succinate	**8.7**	+
Choline	**8.4**	+
2-Oxoglutarate	**6.4**	+
Xanthine	6.3	-
Creatine	**6.2**	+
Alanine	**6.0**	+
Citrate	6.0	-
Glucose	5.9	+
Fumarate	**5.9**	+
Ornithine	5.4	+
Lactate	**5.1**	+
Uridine	**4.7**	-
Serine	4.7	-
Tyrosine	4.1	+
Betaine	2.7	-
Arginine	2.3	-
Glycerol	2.0	-
Dimethylglycine	1.8	+
Phenylalanine	1.6	+
Methionine	1.6	-
Malate	**1.4**	+
Quinolinate	0.8	-
Aconitate	0.7	+
Tryptophan	0.5	-
Proline	0.4	+
Asparagine	0.2	-
Aspartate	0.1	+
ADMA/SDMA	0.1	-
Glycerol 3-phosphate	0.0	-

Table S1: Metabolite weights in a linear classifier of RCD patients and controls based on the cell culture-derived metabolic profile.

The 29 metabolites are ordered by decreasing weight. In bold: the 9 metabolites which were significant in both media and plasma and concordant between the two systems (Figure 3). For each metabolite, its median weight over 100 cross-validation rounds is presented.